

Rates of contraction of posterior distributions based on p-exponential priors

Article (Accepted Version)

Agapiou, Sergios, Dashti, Masoumeh and Helin, Tapio (2021) Rates of contraction of posterior distributions based on p-exponential priors. *Bernoulli - Journal of the Bernoulli Society (Bernoulli)*, 27 (3). pp. 1616-1642. ISSN 1350-7265

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/94192/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Rates of contraction of posterior distributions based on p -exponential priors

Sergios Agapiou¹, Masoumeh Dashti² and Tapio Helin³

¹*Department of Mathematics and Statistics, University of Cyprus. e-mail: agapiou.sergios@ucy.ac.cy*

²*Department of Mathematics, University of Sussex. e-mail: m.dashti@sussex.ac.uk*

³*School of Engineering Science, LUT University. e-mail: tapio.helin@lut.fi*

Abstract: We consider a family of infinite dimensional product measures with tails between Gaussian and exponential, which we call p -exponential measures. We study their measure-theoretic properties and in particular their concentration. Our findings are used to develop a general contraction theory of posterior distributions on nonparametric models with p -exponential priors in separable Banach parameter spaces. Our approach builds on the general contraction theory for Gaussian process priors in [59], namely we use prior concentration to verify prior mass and entropy conditions sufficient for posterior contraction. However, the specific concentration properties of p -exponential priors lead to a more complex entropy bound which can influence negatively the obtained rate of contraction, depending on the topology of the parameter space. Subject to the more complex entropy bound, we show that the rate of contraction depends on the position of the true parameter relative to a certain Banach space associated to p -exponential measures and on the small ball probabilities of these measures. For example, we apply our theory in the white noise model under Besov regularity of the truth and obtain minimax rates of contraction using (rescaled) α -regular p -exponential priors. In particular, our results suggest that when interested in spatially inhomogeneous unknown functions, in terms of posterior contraction, it is preferable to use Laplace rather than Gaussian priors.

MSC 2010 subject classifications: Primary 62G20; secondary 62G05, 60G50.

Keywords and phrases: Bayesian nonparametric inference, non-Gaussian priors, concentration of measure.

1. Introduction

Gaussian processes are routinely used as priors in many nonparametric inference problems, for example in spline smoothing [33], density estimation [39], nonparametric regression [50], inverse problems [52] and drift estimation of diffusions [42]. At the same time, there is a growing number of problems for which it is preferable to utilize heavier-tailed priors, while maintaining the favourable convexity properties offered by the Gaussian distribution. For example, priors

constructed using infinite products of Laplace distributions are extensively used in the literature of Bayesian inverse problems in the form of Besov-space priors with integrability parameter $p = 1$ [37, 17, 36, 29, 1]. Such priors on the one hand have attractive sparsity-promoting properties at the level of maximum a posteriori estimates [36, 1], and on the other hand are logarithmically concave, thus computationally and analytically tractable. Besov-space priors are defined through expansions in a wavelet basis and for $p = 1$ use ℓ_1 -type penalization on the corresponding coefficients, an idea widely-used in the statistical literature [10, 11, 18, 19, 20, 30].

The study of the asymptotic performance of posterior distributions in the infinitely-informative data-limit, under the frequentist assumption that the available data is generated from an underlying fixed value of the unknown, has received great attention in the last two decades. In particular, there has been enormous progress in the study of rates of posterior contraction, that is the concentration rates of posterior distributions around the underlying value of the unknown. The works of Ghosal and van der Vaart [23] and Shen and Wasserman [48] for independent and identically distributed (i.i.d.) observations, together with the work of Ghosal and van der Vaart [24] for non-i.i.d. observations, paved the way for a comprehensive theory for rates of posterior contraction under general assumptions on the prior and model.

For Gaussian priors, posterior contraction has been vigorously studied aided by the available very deep understanding of Gaussian processes; see for example [60] for a presentation of the relevant elements of Gaussian process theory. Of great importance in this context, has been the work of van der Vaart and van Zanten [59], who studied general posterior contraction based on the concentration properties of the Gaussian prior. In particular, they showed that the rate of contraction depends on the position of the true parameter underlying the data relative to the reproducing kernel Hilbert space and the centered small ball probabilities of the Gaussian prior. An incomplete list of other contributions which advanced the theory of posterior contraction under Gaussian priors in several models, often using mixtures of Gaussian processes to achieve adaptation, includes [5, 12, 26, 35, 43, 53, 58, 61, 62]. See also the recent books [25, 27].

On the contrary, the frequentist asymptotic performance of posterior distributions arising from infinite-dimensional Laplace-type priors is much less understood. In particular, there is no general theory for posterior contraction and the only applicable contraction result we are aware of, refers to undersmoothing product priors in the white noise model [13, Corollary 3]. Of some relevance are existing posterior contraction results under sieve priors, which include randomly truncated products of exponential distributions [3, 44]. For such priors a mechanism for choosing the truncation point is necessary either using a hyperprior or with an empirical Bayes procedure. Also relevant, is the result of [14, Section 3] for the independent and identically distributed product Laplace prior in the sparse Gaussian sequence model setting.

In this work, we consider a class of infinite-dimensional priors spanning between Gaussian and Laplace product priors. We call such priors *p-exponential*,

with $p \in [1, 2]$ reflecting the tail behaviour, where $p = 2$ corresponds to Gaussian and $p = 1$ to exponential tails. Our aim is twofold: first, to develop the relevant measure theory for these priors and to study their concentration properties and second, to study posterior contraction for general models based on prior concentration, analogously to the Gaussian contraction theory in [59].

1.1. General posterior contraction theory

Consider the problem of inferring an unknown parameter $\theta \in \Theta$ from observations $X^{(n)}$ drawn from distributions $P_\theta^{(n)}$, where $n \rightarrow \infty$ corresponds to the infinitely-informative data-limit. We put a prior Π on θ and aim to study the frequentist asymptotic properties of the resulting posterior distribution on θ after observing $X^{(n)}$, $\Pi_n(\cdot|X^{(n)})$. In particular, we make the frequentist assumption that the available observations have been generated from a fixed underlying true parameter $\theta_0 \in \Theta$, and we are interested in investigating the concentration rate of the posterior distribution around the truth in the limit $n \rightarrow \infty$. We say that the posterior distribution contracts with a rate ϵ_n at θ_0 with respect to a metric d on Θ , if $\Pi_n(\theta : d(\theta, \theta_0) \geq M_n \epsilon_n | X^{(n)}) \rightarrow 0$ in $P_{\theta_0}^{(n)}$ -probability, for every $M_n \rightarrow \infty$.

Posterior contraction in this general-prior and general-model setup, has been studied by Ghosal and van der Vaart in [24]. Given a model and distance d , assuming that there exist exponentially powerful tests for separating θ_0 from d -balls at a certain distance from it, they derived conditions on the prior securing that an ϵ_n is a rate of contraction around θ_0 with respect to d : the prior needs to put sufficient mass around the true θ_0 and almost all its mass on sets of bounded complexity. These conditions are expressed via norms and discrepancies which are relevant to the statistical setting of interest. In particular, they involve both neighbourhoods of θ_0 expressed via the metric d , as well as neighbourhoods of $P_{\theta_0}^{(n)}$ expressed via Kullback-Leibler divergence and variations. For a comprehensive and up to date treatment see [25, Chapter 8].

1.2. Gaussian concentration and posterior contraction

We briefly describe the posterior contraction theory for Gaussian priors of van der Vaart and van Zanten [59], which relies on a good understanding of the concentration properties of Gaussian measures; see also [25, Chapter 11].

Let $(X, \|\cdot\|)$ be a separable Banach space and let μ be a centered Gaussian prior in X . Denote by H the reproducing kernel Hilbert space (RKHS) of μ , with corresponding norm $\|\cdot\|_H$. Moreover, denote by B_X the closed unit ball of X centered at the origin. The concentration properties of μ at a point w in the topological support of μ , $\text{supp}(\mu) = \overline{H}^{\|\cdot\|_X}$, were shown to be captured by the concentration function

$$\varphi_w(\epsilon) = \inf_{h \in H: \|h-w\| \leq \epsilon} \frac{1}{2} \|h\|_H^2 - \log \mu(\epsilon B_X), \quad \epsilon > 0. \quad (1)$$

For $w = 0$, the first term vanishes and the concentration function measures the probability of centered balls of size ϵ in X . The idea is that for nonzero $w \in \text{supp}(\mu)$, the concentration function measures the probability of balls of radius ϵ centered at w , with the first term measuring the loss of probability due to shifting from centered to noncentered balls; this is made precise by the bounds in [60, Lemma 5.3].

Using the above interpretation of the concentration function, together with a concentration inequality due to Borell, [9, Theorem 3.1], van der Vaart and van Zanten showed in [59, Theorem 2.1] that for a $w_0 \in \text{supp}(\mu)$, if ϵ_n satisfies

$$\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2, \quad (2)$$

then the prior puts a certain minimum mass in ϵ_n balls in X around the w_0 and it is possible to find $\Theta_n \subset X$ which contains the bulk of the prior mass and has exponentially bounded complexity. These assertions point to the conditions of general-model general-prior results discussed in the previous subsection, see for example [25, Theorem 8.9 and Theorem 8.19]. However unlike the conditions of these general results which involve statistically relevant norms and discrepancies, the assertions of [59, Theorem 2.1] are expressed purely in the Banach space norm. To bridge this gap and indeed prove that ϵ_n is a posterior contraction rate in specific statistical settings, one needs to relate the statistically relevant quantities appearing in general-model general-prior results to the Banach space norm.

In a range of models, this reconciliatory work has been done in [24] in the general-prior context, and there exist general-prior contraction theorems with assumptions purely expressed in the Banach space norm [24]; for example see [25, Theorem 8.31] in the white noise model, or [25, Theorem 8.26] in the normal fixed-design regression setting. In other models such as density estimation or nonparametric binary classification, the reconciliatory work has been done in the context of Gaussian priors in [59], see [59, Lemmas 3.1 and 3.2] respectively, and no general-prior theorems were explicitly formulated. We stress that the reconciliatory work for these models is not explicit to Gaussian priors, thus the proofs of all the Gaussian contraction results found in [59, Section 3], can be easily used to get contraction results for priors for which analogous results to [59, Theorem 2.1] hold.

1.3. Our contribution

In the present paper we consider parameter spaces X which are separable Banach and which possess a Schauder basis. We use the Schauder basis to construct p -exponential measures in X , by identifying them to infinite products of independent univariate p -exponential distributions. Our main contribution is that we generalize the aforementioned Gaussian general contraction theorem [59, Theorem 2.1] to p -exponential measures, and to achieve this we develop the necessary concentration theory for p -exponential measures. The obtained general contraction result enables the study of contraction rates of posterior distributions based

on p -exponential priors, in a range of standard nonparametric statistical models. A brief summary of the paper is as follows.

In Section 2, we introduce p -exponential measures in X and study their properties relating to convexity, equivalence and singularity under translations, topological support and ultimately concentration. We find that the concentration of a p -exponential measure at a point w in its support, depends on the position of w relative to a Banach space, rather than relative to a Hilbert space as was the case for Gaussian measures. We define the corresponding concentration function $\varphi_w(\cdot)$ and show in Theorem 2.13 that it has a similar interpretation to the Gaussian concentration function. In Proposition 2.15, we derive a concentration inequality for p -exponential measures, which follows from Talagrand's work in [54] and, although substantially more intricate, is analogous to the aforementioned Gaussian concentration inequality [9, Theorem 3.1] used for studying contraction in [59].

In Section 3 we use the interpretation of the concentration function, together with the available concentration inequality to generalize the Gaussian contraction result [59, Theorem 2.1] to p -exponential measures in Theorem 3.1, which is the main result of this paper. Since the concentration properties of p -exponential measures are more intricate, we get a more complicated complexity bound compared to the Gaussian case.

In Section 4, we present posterior contraction results for general p -exponential measures in two standard statistical models: the white noise model and density estimation. These results follow immediately from Theorem 3.1, as discussed at the end of the last subsection.

In Section 5 we consider α -regular p -exponential priors in separable Hilbert spaces, study bounds on the corresponding concentration function for Besov-type regularity of the truth and compute posterior contraction rates in the white noise model, with L_2 loss. In this case, the complexity bound in Theorem 3.1, which is more complicated for p -exponential priors compared to Gaussian priors, does not affect the rates. Our bounds are particularly interesting for Besov spaces of spatially inhomogeneous functions, that is for Besov integrability parameter $q < 2$. In this case, Gaussian priors appear to be suboptimal, more specifically to be limited by the minimax rate over linear estimators, which is slower than the minimax rate [21, Theorem 1]. On the other hand, p -exponential priors with $p < 2$ can do better than the linear minimax rate, see Theorem 5.5 and Remark 5.7. Furthermore, we can achieve the minimax rate using rescaled undersmoothing p -exponential priors for $p = q$, or the minimax rate up to lower order logarithmic terms using rescaled undersmoothing p -exponential priors with $p < q$; see Theorem 5.9 and Remark 5.10. To our knowledge, this is the first occurrence of a prior achieving the minimax rate over such Besov spaces in the literature. Although we only show upper bounds on the rate of contraction, our results indicate that when interested in spatially inhomogeneous unknowns, in terms of posterior contraction rates, it is beneficial to use Laplace rather than Gaussian priors.

In Section 6 we consider α -regular p -exponential priors constructed via wavelet expansions in the space of continuous functions on the unit interval, $C[0, 1]$, and

study bounds on the corresponding concentration function in the supremum norm, under Hölder-type regularity of the truth. To this end, we prove new centered small ball probability bounds in the supremum norm for p -exponential measures, see Proposition 6.3. We then compute posterior contraction rates in density estimation, with Hellinger-distance loss. In this case, the rates are affected by the more complicated complexity bound in Theorem 3.1 and appear to be suboptimal, see Theorem 6.7.

The proofs of our results are contained in the supplementary material below, together with some complementary technical results and discussions.

1.4. Notation

We denote by \mathbb{R}^∞ the space of all real sequences and by $\mathcal{B}(\mathbb{R}^\infty)$ the Borel σ -algebra with respect to the product topology. We denote by ℓ_p the space of p -summable real sequences. The space of square integrable real functions on the unit interval is denoted by $L_2[0, 1]$, while $C[0, 1]$ is the space of continuous real functions on the unit interval with the supremum norm. For $s > 0$, we use $C^s = C^s[0, 1]$ to denote the space of s -Hölder real functions on the unit interval. For a normed space $(Y, \|\cdot\|_Y)$, we denote by B_Y the closed unit ball in Y . The notation $N(\epsilon, A, d)$ is used for the ϵ -covering number of a subset A of a metric space with metric d , that is the minimum number of balls of radius ϵ with respect to d which are needed to cover the set A . For two positive sequences $(a_n), (b_n)$, $a_n \asymp b_n$ means a_n/b_n is bounded away from zero and infinity, while $a_n \lesssim b_n$ means that a_n/b_n is bounded.

2. p -exponential measures and their properties

In this section we introduce p -exponential measures and study some of their properties. In particular, we discuss their convexity, behaviour under translations, topological support and concentration properties.

2.1. p -exponential measures

Definition 2.1. Let $\gamma = (\gamma_\ell)_{\ell \in \mathbb{N}}$ be a deterministic decaying sequence of positive real numbers and let $\xi_\ell, \ell \in \mathbb{N}$, be independent and identically distributed real random variables with probability density function $f_p(x) \propto \exp(-\frac{|x|^p}{p})$, $x \in \mathbb{R}$ for $p \in [1, 2]$. We define the probability measure μ on the measurable space $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ to be the law of the sequence $(\gamma_\ell \xi_\ell)_{\ell \in \mathbb{N}}$ and call it a p -exponential measure with scaling sequence γ .

In the following we will often suppress the dependence on γ and call μ a p -exponential measure. For $p = 1$ and $p = 2$ we get centered Laplace and centered Gaussian measures respectively, both in sequence space. While we restrict p between 1 and 2, many of the results in this section as for example the ones in the following subsection on convexity, clearly hold in greater generality and

in particular for $p \geq 1$. However, our treatment on the concentration of p -exponential measures in Subsection 2.4, is explicit to $p \in [1, 2]$.

Depending on the decay properties of γ , draws from μ almost surely belong to certain subspaces of \mathbb{R}^∞ . For example, $\gamma \in \ell_2$ if and only if $\mu(\ell_2) = 1$, see [2, Lemma S.M.1.2]. Lemma 5.2 in Section 5 below studies Besov-type regularity of μ , for certain choices of the scaling sequence γ ; this result includes Sobolev-type regularity as a special case.

Any Gaussian random element in a separable Banach space can be identified with a Gaussian product measure as above with $p = 2$, for example using the Karhunen-Loeve expansion [27, Theorem 2.6.10]. Likewise, a p -exponential measure can be identified naturally with a measure on a separable Banach space X , provided X possesses a Schauder basis, which can be normalized or not.

Definition 2.2. *Let $(X, \|\cdot\|_X)$ be a separable Banach space. A Schauder basis is a sequence $\{\psi_\ell\} \subset X$, such that for every $u \in X$, there exists a unique real sequence $(u_\ell)_{\ell \in \mathbb{N}}$, so that*

$$u = \sum_{\ell=1}^{\infty} u_\ell \psi_\ell,$$

where the convergence is with respect to $\|\cdot\|_X$.

For example, if $\gamma \in \ell_2$, a p -exponential measure can be identified with a measure on a subspace of the space of square integrable functions on the unit interval, $X = L_2[0, 1]$, via the random series expansion

$$u(x) = \sum_{\ell=1}^{\infty} \gamma_\ell \xi_\ell \psi_\ell(x), \quad (3)$$

where $\{\psi_\ell\}$ is an orthonormal basis in $L_2[0, 1]$. It can also be identified with a measure on the space of continuous functions on the unit interval, $X = C[0, 1]$, using a similar random series expansion, where $\{\psi_\ell\}$ is a Schauder basis in $C[0, 1]$; see Section 6 below.

In the general separable Banach space setting, we also have that depending on the speed of decay of the scaling sequence γ , draws from a p -exponential measure almost surely belong to subspaces of X . If X is a function-space, these subspaces correspond to a form of higher regularity. We stress here, that such function-space regularity is not solely linked to the speed of decay of γ , but also depends on the scaling and regularity of the Schauder basis $\{\psi_\ell\}$. For example, one can study the Hölder regularity of draws using the Kolmogorov Continuity Test. See [17, Corollary 5] for a result under general conditions on the Schauder basis and scaling sequence, or Proposition 6.1 in Section 6 below for a result under more specific conditions.

While developing our posterior contraction theory for p -exponential priors below, we will use the sequence space or the general separable Banach space representation of the measure μ interchangeably. The particular random series expansion representation, and specifically the choice of the Schauder basis, will become relevant through the concentration function when actually computing the contraction rate in specific settings with specific priors in Sections 5 and 6.

2.2. Convexity

We next study the convexity properties of p -exponential measures. The convexity of measures in infinite dimensional spaces has been extensively studied in [8].

Proposition 2.3. *A p -exponential measure μ is logarithmically-concave. That is, for any measurable sets $A, B \in \mathcal{B}(\mathbb{R}^\infty)$ and any $s \in [0, 1]$ it holds*

$$\mu(sA + (1-s)B) \geq \mu(A)^s \cdot \mu(B)^{1-s}.$$

This is a straightforward result based on [8]. A proof, done for a specific type of choice of γ without loss of generality, can be found in [1, Lemma 3.4]. Logarithmic concavity is a very strong property which for example implies unimodality, see [1, Section 2]. An immediate consequence is the following inequality called Anderson's inequality, implied by [8, Theorem 6.1], which holds since we consider centered measures.

Proposition 2.4. *Let μ be a p -exponential measure. For any closed, symmetric and convex set $A \subset \mathbb{R}^\infty$, we have*

$$\mu(A+x) \leq \mu(A), \forall x \in \mathbb{R}^\infty.$$

Logarithmic-concavity also implies the following zero-one law, see [8, Theorem 4.1].

Proposition 2.5. *Let μ be a p -exponential measure. Then for any linear subspace $V \subset \mathbb{R}^\infty$ we have that $\mu(V) = 0$ or 1.*

2.3. Absolute continuity

We next consider the equivalence or singularity of a p -exponential measure to its translations.

Definition 2.6. *For a measure ν on a measurable space $(\mathcal{X}, \mathcal{F})$, we define the space of admissible shifts $\mathcal{Q} = \mathcal{Q}(\nu)$ to be the subspace of all translations $h \in \mathcal{X}$ such that $\nu_h(\cdot) := \nu(\cdot - h)$ is equivalent to ν as measures.*

The next proposition identifies the space of admissible shifts of the p -exponential measure and provides an expression for the Radon-Nikodym derivative of μ_h with respect to μ , for $h \in \mathcal{Q}(\mu)$. It also shows that the two measures are singular for $h \notin \mathcal{Q}(\mu)$.

Proposition 2.7. *Let μ be a p -exponential measure and let $h \in \mathbb{R}^\infty$. Then μ_h and μ are either equivalent or singular. The space of admissible shifts of μ is*

$$\mathcal{Q} = \mathcal{Q}(\mu) = \{h \in \mathbb{R}^\infty : \sum_{\ell=1}^{\infty} h_\ell^2 \gamma_\ell^{-2} < \infty\}.$$

In particular, it is a separable Hilbert space with norm

$$\|h\|_{\mathcal{Q}} = \left(\sum_{\ell=1}^{\infty} h_{\ell}^2 \gamma_{\ell}^{-2} \right)^{\frac{1}{2}}, \quad \forall h \in \mathcal{Q}.$$

Furthermore, for $h \in \mathcal{Q}(\mu)$,

$$\frac{d\mu_h}{d\mu}(u) = \lim_{N \rightarrow \infty} \exp \left(\frac{1}{p} \sum_{\ell=1}^N \left(\left| \frac{u_{\ell}}{\gamma_{\ell}} \right|^p - \left| \frac{h_{\ell} - u_{\ell}}{\gamma_{\ell}} \right|^p \right) \right) \quad \text{in } L^1(\mathbb{R}^{\infty}, \mu).$$

The last result is an immediate application of a more general result valid for scaled independent products of univariate distributions with finite Fisher information and everywhere positive density, see Proposition E.1 in the supplement below.

Even though the Radon-Nikodym derivative between a centered and a translated p -exponential measure involves weighted ℓ_p -type terms, the space of admissible shifts, even for $p \neq 2$, is a weighted ℓ_2 space, that is a Hilbert space. Furthermore, it is straightforward to check that $\mu(\mathcal{Q}(\mu)) = 0$. Indeed, for u drawn from a p -exponential measure we have $\|u\|_{\mathcal{Q}}^2 = \sum_{\ell=1}^{\infty} \xi_{\ell}^2$ which is almost surely infinite by the law of large numbers.

Motivated by the exponent of the Radon-Nikodym derivative above, we define the following subspace.

Definition 2.8. For a p -exponential measure μ , we define the separable Banach space

$$\mathcal{Z} = \mathcal{Z}(\mu) = \{h \in \mathbb{R}^{\infty} : \sum_{\ell=1}^{\infty} |h_{\ell}|^p \gamma_{\ell}^{-p} < \infty\},$$

with norm

$$\|h\|_{\mathcal{Z}} = \left(\sum_{\ell=1}^{\infty} |h_{\ell}|^p \gamma_{\ell}^{-p} \right)^{\frac{1}{p}}, \quad \forall h \in \mathcal{Z}.$$

The space $\mathcal{Z}(\mu)$ is a weighted ℓ_p space which, since $p \in [1, 2]$ and γ_{ℓ} is a decaying sequence, is continuously embedded in $\mathcal{Q}(\mu)$. Clearly we also have that $\mu(\mathcal{Z}(\mu)) = 0$.

When working in a separable Banach space X possessing a Schauder basis, the subspaces $\mathcal{Q}(\mu) \subset \mathbb{R}^{\infty}$ and $\mathcal{Z}(\mu) \subset \mathbb{R}^{\infty}$ are naturally identified with subspaces of X . If X is a function-space, then \mathcal{Z} and \mathcal{Q} correspond to subspaces of functions of higher regularity. In the Gaussian case $p = 2$, we have that \mathcal{Q} and \mathcal{Z} are identified with the RKHS [25, Section I.6], but in general the two spaces differ and have different roles.

For Gaussian measures, the RKHS is compactly embedded in any separable Banach space X of full measure, [27, Proposition 2.6.9]. The next proposition generalizes this statement for p -exponential measures. It follows from [7, Theorem 5.1.6.], which holds for general Radon measures on locally convex spaces.

Proposition 2.9. Let μ be a p -exponential measure on a separable Banach space X with a Schauder basis. The space of admissible shifts $\mathcal{Q}(\mu)$ is compactly embedded into X . As a consequence, $\mathcal{Z}(\mu)$ is also compactly embedded into X .

2.4. Support and concentration

In this subsection, $(X, \|\cdot\|_X)$ is a separable Banach space possessing a Schauder basis and μ is a p -exponential measure on X , $\mu(X) = 1$, defined by randomizing the coefficients of random series expansions in the Schauder basis as explained in Subsection 2.1. For a Gaussian measure on X , it is known that its topological support is the closure of the RKHS in X , [27, Corollary 2.6.17]. We next show an analogous result for p -exponential measures. Since γ is a sequence of positive scalings, p -exponential measures are non-degenerate, that is their support is the whole space X .

Proposition 2.10. *Let μ be a p -exponential measure on X . Then*

$$X = \text{supp}(\mu) = \overline{\mathcal{Q}}^{\|\cdot\|_X} = \overline{\mathcal{Z}}^{\|\cdot\|_X},$$

where $\overline{\mathcal{Q}}^{\|\cdot\|_X}, \overline{\mathcal{Z}}^{\|\cdot\|_X}$ denote the closures in the norm $\|\cdot\|_X$ of the spaces \mathcal{Q}, \mathcal{Z} , respectively.

The role of the subspace \mathcal{Z} is revealed in the next two results, which study the probability of non-centered balls in X relative to the probability of centered ones, under a p -exponential measure. Proposition 2.4, showed that for fixed radius $\epsilon > 0$, there is a loss of probability when shifting from centered to non-centered balls. In the next proposition we prove a lower bound on the loss of probability when the shift is in the space \mathcal{Z} .

Proposition 2.11. *Let μ be a p -exponential measure on X . Then for $h \in \mathcal{Z}$ and any $\epsilon > 0$, we have*

$$\mu(\epsilon B_X + h) \geq e^{-\frac{1}{p} \|h\|_{\mathcal{Z}}^p} \mu(\epsilon B_X).$$

Our proof relies on certain properties of the function $|\cdot|^p, p \in [1, 2]$ appearing in the exponent of the Radon-Nikodym derivative $d\mu_h/d\mu$ in Proposition 2.7, namely its symmetry together with its convexity and the concavity of its derivative on the positive semi-axis.

For $p = 2$ we recover the Gaussian result [60, Lemma 5.2]. For $p \neq 2$, the loss of probability is exponential in the \mathcal{Z} -norm and not in the Hilbert space norm of the space of admissible shifts \mathcal{Q} . As we will see in the next section, this adds a degree of difficulty to the study of posterior contraction for p -exponential priors. Due to the form of the Radon-Nikodym derivative in Proposition 2.7, the last result is not surprising. In particular, it is consistent with the form of the Onsager-Machlup functional, that is the functional giving the most probable paths, for Besov-space measures with $p = 1$ in [1, Theorem 3.9].

We next extend the last lower bound to centers that are not necessarily in \mathcal{Z} , using approximation. We restrict to centers in the topological support of μ , X , since otherwise a small enough ball around w has zero probability. As in the Gaussian case, see (1) in Subsection 1.2, we define the concentration function of a p -exponential measure.

Definition 2.12. Let $w \in X$. We define the concentration function of the p -exponential measure μ on X to be

$$\varphi_w(\epsilon) = \inf_{h \in \mathcal{Z}: \|h-w\|_X \leq \epsilon} \frac{1}{p} \|h\|_{\mathcal{Z}}^p - \log \mu(\epsilon B_X).$$

The first term relates to approximation of the center $w \in X = \overline{\mathcal{Z}}^{\|\cdot\|_X} = \overline{\mathcal{Q}}^{\|\cdot\|_X}$ by elements of the space \mathcal{Z} . Unlike the Gaussian case and consistently with Proposition 2.10, for $p \neq 2$ this approximation does not take place in a Hilbert space. For any $w \in X$, since the \mathcal{Z} -norm is convex and p -exponential measures are logarithmically-concave and non-degenerate, the concentration function is a strictly decreasing and convex function on the positive semi-axis. This follows very similarly to the Gaussian case see [12, Lemma 3] or the more readily adaptable [25, Lemma I.26]. In particular, the concentration function is continuous and blows-up as $\epsilon \rightarrow 0$. Depending on the position of $w \in X$ relative to the space \mathcal{Z} , the blow-up rate is determined by the first or second term. For example, if $w \in \mathcal{Z}$ the first term remains bounded and only the second term blows-up.

The interpretation of the concentration function is similar to the Gaussian case. For $w = 0$, the first term is zero and $\varphi_0(\epsilon)$ measures the probability with respect to μ of a centered ball of radius ϵ in X . For $w \in X \setminus \{0\}$, the next theorem shows that the concentration function gives a lower bound on the probability of a ball of radius ϵ in X around w , with the first term measuring the loss of probability due to moving the ball away from the origin.

Theorem 2.13. For any $w \in X$ we have that

$$-\log \mu(w + \epsilon B_X) \leq \varphi_w(\epsilon/2), \quad \forall \epsilon > 0.$$

The proof of the last theorem is very similar to the first part of the proof of [60, Lemma 5.3]. It follows from Proposition 2.11 using the triangle inequality and approximation of $w \in X$ in \mathcal{Z} .

Remark 2.14. In the Gaussian case, the concentration function yields both an upper and a lower bound on the probability of small balls around a $w \in X$ [60, Lemma 5.3]. While the last theorem achieves a lower bound, it would be interesting to also prove an upper bound in the p -exponential case. However, the lack of inner product structure in the Radon-Nikodym derivative between a centered p -exponential measure and its translation makes this task considerably harder.

The following inequality generalizes Borell's inequality which studies the concentration of Gaussian measures, [9, Theorem 3.1]. It is based on a sharp two level concentration inequality due to Michel Talagrand [54, Theorem 2.4].

Proposition 2.15. Let μ be a p -exponential measure in X . Recall that $B_{\mathcal{Z}}$ and $B_{\mathcal{Q}}$ denote the closed unit balls in the spaces \mathcal{Z} and \mathcal{Q} , respectively. Then there exists a constant $K > 0$ depending only on p , such that for any set $A \in \mathcal{B}(X)$ and any $r > 0$ it holds

$$\mu(A + r^{\frac{p}{2}} B_{\mathcal{Q}} + r B_{\mathcal{Z}}) \geq 1 - \frac{1}{\mu(A)} \exp\left(-\frac{r^p}{K}\right). \quad (4)$$

Letting $A = \epsilon B_X$ for a fixed small $\epsilon > 0$, the last inequality implies that while both \mathcal{Z}, \mathcal{Q} are null sets of μ , the bulk of the mass of μ is contained in a small ϵ -cushion in X around the sum of a ball of radius r in \mathcal{Z} and a ball of radius $r^{\frac{p}{2}}$ in \mathcal{Q} , for r large. This interpretation is similar to the one for Borell's inequality presented in the discussion after [25, Proposition 11.17], which is simpler since in the Gaussian case $\mathcal{Z} = \mathcal{Q}$.

Remark 2.16. *Borell's inequality [9, Theorem 3.1] for Gaussian measures has the form of a stronger isoperimetric inequality, which in turn implies the concentration inequality (4) in the case $p = 2$ and $\mathcal{Q} = \mathcal{Z}$. Using results in isoperimetry for finite independent products of standard univariate p -exponential distributions [45], together with the techniques in [9] to pass from finite to infinite dimensions, one can show that there exists $K = K(p) > 0$ such that for any $A \in \mathcal{B}(X)$ it holds*

$$\mu(A + rB_{\mathcal{Q}}) \geq F_p(F_p^{-1}(\mu(A)) + Kr),$$

where F_p is the cumulative distribution function of the univariate standard p -exponential distribution. The concentration inequality implied by the above inequality has the form

$$\mu(A + rB_{\mathcal{Q}}) \geq 1 - \frac{1}{\mu(A)} \exp\left(-\frac{r^p}{K}\right),$$

and for $p \in [1, 2)$ is strictly weaker than the one in Proposition 2.15, since it involves balls of radius r in the space \mathcal{Q} which strictly contains \mathcal{Z} .

3. General contraction theorem for p -exponential priors

We next state our general contraction result for p -exponential priors in a separable Banach space X possessing a Schauder basis, which generalizes the Gaussian contraction result [59, Theorem 2.1]. It shows that for a p -exponential prior and a $w_0 \in X$, if ϵ_n is such that the blow-up rate of the concentration function satisfies

$$\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2, \tag{5}$$

then there exist sets $X_n \subset X$ of bounded complexity containing the bulk of the prior mass, and the prior puts sufficient mass around w_0 . These assertions are in accordance with the requirements of results giving upper bounds on the contraction rate at w_0 for general priors, see the discussion in Subsection 1.2 and the results in Section 4 below.

To prove our contraction result, we follow the techniques of the proof of the Gaussian result [59, Theorem 2.1], which is based on Borell's inequality [9, Theorem 3.1] together with the concentration function and its relation to lower bounds on the probability of shifted small balls [60, Lemma 5.3]. However, the situation for p -exponential priors is more complicated, due to the intricate form of the available concentration inequality in Proposition 2.15. In particular, due to the fact that for $p \in [1, 2)$, the concentration inequality (4) involves both balls

in \mathcal{Q} and balls in \mathcal{Z} , while the decentering result in Proposition 2.11 refers to elements in \mathcal{Z} , in order to prove the complexity bound we need to approximate elements in \mathcal{Q} by elements in \mathcal{Z} . To this end we let $f, g : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ be two respectively non-decreasing and non-increasing functions, such that for $\epsilon, a > 0$ and for any $h \in aB_{\mathcal{Q}}$ it holds

$$\inf_{x \in \mathcal{Z} : \|x-h\|_X \leq \epsilon} \|x\|_{\mathcal{Z}}^p \leq f(a)g(\epsilon)^{1-\frac{p}{2}} \quad (6)$$

and as $a \rightarrow \infty$, $f(a)$ grows at most polynomially to infinity. For $p = 2$, since $\mathcal{Q} = \mathcal{Z}$, we can choose $f(a) = a^2$ while g is redundant. For $p \in [1, 2)$, since $\mathcal{Z} \subsetneq \mathcal{Q}$, g needs to satisfy $g(\epsilon) \rightarrow \infty$ as $\epsilon \rightarrow 0$. For optimal results we need to choose f and g so that the bound (6) is as tight as possible. As a result of this extra approximation step, we get a more complicated form on the right hand side of the complexity bound, see (7) below, compared to the Gaussian case [59, Theorem 2.1]. Note that the factorization of the right hand side of (6) into the two functions f and g , is not important for the theory, but arises naturally in practice, see Lemmas 5.14 and 6.6 below.

Theorem 3.1. *Let μ be a p -exponential measure with scaling sequence γ in a separable Banach space X with Schauder basis, where $p \in [1, 2]$. Let $W \sim \mu$. Fix $f, g : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$, as in (6) above and let $w_0 \in X$.*

Assume $\epsilon_n > 0$ such that $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$, where $n\epsilon_n^2 \gtrsim 1$. Then for any $C > 1$, there exists a measurable set $X_n \subset X$ and a constant $R > 0$ depending on C, p and f , such that

$$\log N(4\epsilon_n, X_n, \|\cdot\|_X) \leq R(n\epsilon_n^2 \vee f(n^{\frac{1}{2}}\epsilon_n)g(\epsilon_n)^{1-\frac{p}{2}}), \quad (7)$$

$$\mathbb{P}(W \notin X_n) \leq \exp(-Cn\epsilon_n^2), \quad (8)$$

$$\mathbb{P}(\|W - w_0\|_X < 2\epsilon_n) \geq \exp(-n\epsilon_n^2). \quad (9)$$

The difference between the assertions of the above theorem compared to the Gaussian result [59, Theorem 2.1], is the right hand side in the complexity bound (7), which is potentially larger than $n\epsilon_n^2$, depending on which of the two terms dominates in the maximum asymptotically as $\epsilon_n \rightarrow 0$. In the Gaussian case $p = 2$, the right hand side in (7) becomes $n\epsilon_n^2$ and we recover [59, Theorem 2.1]. For $p \in [1, 2)$, depending on the norm in the parameter space X , we have a different form of the tightest functions f and g that we can verify to satisfy (6). If the quality of approximation in $\|\cdot\|_X$ of elements in \mathcal{Q} by elements in \mathcal{Z} is not sufficiently good, the right hand side in (7) can be dominated by the second term and in this case the complexity bound is not in accordance with the corresponding complexity bound in general prior contraction results like [25, Theorem 8.9 and 8.19]. The next corollary handles such situations.

Corollary 3.2. *Under the assumptions of Theorem 3.1, let $\tilde{\epsilon}_n > 0$ be such that*

$$f(n^{\frac{1}{2}}\epsilon_n)g(\epsilon_n)^{1-\frac{p}{2}} \lesssim n\tilde{\epsilon}_n^2. \quad (10)$$

Then for any $C > 1$, there exists a measurable set $X_n \subset X$ and a constant $R > 0$, such that

$$\log N(4(\epsilon_n \vee \tilde{\epsilon}_n), X_n, \|\cdot\|_X) \leq Rn(\epsilon_n \vee \tilde{\epsilon}_n)^2, \quad (11)$$

$$\mathbb{P}(W \notin X_n) \leq \exp(-Cn\epsilon_n^2), \quad (12)$$

$$\mathbb{P}(\|W - w_0\|_X < 2\epsilon_n) \geq \exp(-n\epsilon_n^2). \quad (13)$$

The corollary follows immediately from Theorem 3.1, since taking a larger ϵ_n makes the left hand side of the complexity bound (7) smaller and the right hand side larger.

In settings for which (10) is satisfied with $\tilde{\epsilon}_n = \epsilon_n$, for ϵ_n the fastest rate solving (5), we can apply the corollary and the resulting three assertions are in accordance with the general contraction results which show that this ϵ_n is an upper bound on the contraction rate. We will see in Section 5, that this is the case in separable Hilbert space settings for α -regular p -exponential priors and under Besov-type regularity of w_0 . In this situation the intuition about the contraction rate is similar to the Gaussian case, the only difference being that the RKHS is replaced by the Banach space \mathcal{Z} . We refer to the discussion in [25, Section 11.3] which we adapt here to p -exponential priors: the rate of contraction is up to constants the maximum of the minimal solution to the small ball inequality

$$-\log \mu(\epsilon_n B_X) \leq n\epsilon_n^2$$

and the minimal solution to the approximation inequality

$$\inf_{h \in \mathcal{Z}: \|h - w\|_X \leq \epsilon_n} \|h\|_{\mathcal{Z}}^p \leq n\epsilon_n^2.$$

The first inequality depends only on the prior, showing that priors that put little mass around the origin give slow rates independently of w_0 . The second inequality depends on both the prior and the true w_0 and relates to the loss of probability mass in small balls around w_0 compared to centered small balls. It shows that even if the prior puts a lot of mass around the origin, it is still possible to give a slow rate at a w_0 , depending on the positioning of w_0 relative to the Banach space \mathcal{Z} .

On the other hand in settings for which (10) is only satisfied for $\tilde{\epsilon}_n$ a sequence decaying more slowly than the fastest rate ϵ_n solving (5), the resulting three assertions of the corollary are only in accordance with the general contraction result in the independent and identically distributed data case [25, Theorem 8.9], which shows that the slower rate $\tilde{\epsilon}_n$ is an upper bound on the contraction rate. We will see in Section 6 that such issues arise for α -regular p -exponential priors in $C[0, 1]$, defined via wavelet bases. In this case the intuition regarding the rates of contraction is obfuscated.

Remark 3.3. For Gaussian priors, the availability of an upper bound on the probability of small balls around an element $w \in X$ in terms of the concentration

function, enabled the study of lower bounds on posterior contraction rates in [12]. Such an upper bound remains open for p -exponential priors with $p \neq 2$, see remark 2.14, hence the use of the techniques of [12] to similarly obtain lower bounds on posterior contraction rates in this case is precluded.

4. Posterior contraction for specific models

We next use the results of the preceding section to study posterior contraction for general p -exponential priors in specific nonparametric statistical settings. Indeed, the assertions of Theorem 3.1 and Corollary 3.2, point to the assumptions of the well known general model and general prior posterior contraction rate results [25, Theorem 8.9 and 8.19]. However, the former results are expressed purely in terms of the Banach space norm of the parameter space, while the latter have conditions relating to statistically relevant norms and discrepancies. As discussed in Subsection 1.2, the necessary reconciliatory work has already been carried out in various standard statistical settings and can be readily used for p -exponential priors in the same way that it was used for Gaussian priors in [59, Section 3].

Note, that compared to the Gaussian contraction results found in [59, Section 3] or [25, Section 11.3], in the formulation of our results we need to take into account the more complicated complexity bound in (7). For reasons of brevity, we only present here contraction results for density estimation and for the white noise model. Results in other models such as binary classification and nonparametric regression follow similarly.

4.1. Density estimation

We consider the estimation of a probability density π relative to a finite measure ν on a measurable space (T, \mathcal{T}) , based on a sample of observations $X_1, \dots, X_n | \pi \stackrel{iid}{\sim} \pi$. Following [25, Section 11.3.1], we construct a prior Π on π by letting

$$\pi(x) = \frac{e^{W(x)}}{\int_T e^{W(y)} d\nu(y)}, \quad x \in T,$$

where W is a draw from a p -exponential measure μ on $L_\infty(T) \cap C(T)$. We require that W is almost surely continuous so that it can be evaluated at $x \in T$ and $\pi(x)$ is well defined. We can define p -exponential priors with continuous and bounded paths, see Section 6 below.

Let $\Pi_n(\cdot | X_1, \dots, X_n)$ be the posterior distributions after observing X_1, \dots, X_n . The following contraction result is a generalization of the Gaussian result [59, Theorem 3.1]. It gives contraction rates in the Hellinger distance $d_H(\cdot, \cdot)$ between two probability densities. The proof is identical to the Gaussian case, once we take into account Corollary 3.2 (see also [25, Theorem 11.21]).

Theorem 4.1. *Let W be a p -exponential random element in a separable Banach subspace of $L_\infty(T)$ possessing a Schauder basis, which is almost surely continuous. Assume $w_0 = \log \pi_0$ belongs to the support of W and denote by P_0^n the corresponding distribution of the vector (X_1, \dots, X_n) . Let ϵ_n satisfying (5) with respect to $\|\cdot\|_{L_\infty}$ and $\tilde{\epsilon}_n$ satisfying (10) where the functions f, g are defined in (6). Then $\Pi_n(\pi : d_H(\pi, \pi_0) > M(\epsilon_n \vee \tilde{\epsilon}_n) | X_1, \dots, X_n) \rightarrow 0$, in P_0^n -probability, for some sufficiently large constant M .*

4.2. White noise model

We study the estimation of a signal $w \in \Theta \subset L_2[0, 1]$, from the observation of a sample path of the stochastic process

$$X_t^{(n)} = \int_0^t w(s) ds + \frac{1}{\sqrt{n}} B_t, \quad t \in [0, 1]$$

where B is standard Brownian motion. Let $P_w^{(n)}$ be the distribution of the sample path $X^{(n)}$ in $C[0, 1]$. As a prior on w we take a p -exponential random element W in $L_2[0, 1]$. We can define such priors using random series expansions for example in an orthonormal basis of $L_2[0, 1]$, see Section 5 below. Observe that by Proposition 2.10, the topological support of such a prior is $L_2[0, 1]$.

We denote by $\Pi_n(\cdot | X^{(n)})$ the posterior on w after observing the sample path $X^{(n)}$. The following posterior contraction result is a generalization of [59, Theorem 3.4]. The proof is identical to the Gaussian case, and follows immediately by combining [25, Theorem 8.31] and Theorem 3.1.

Theorem 4.2. *Let W be a p -exponential random element in $L_2[0, 1]$. Assume that the true value of w is contained in the support of W , $w_0 \in L_2[0, 1]$. Furthermore, assume that ϵ_n satisfies the rate equation $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$ with respect to the $L_2[0, 1]$ -norm and is such that (7) holds with $n\epsilon_n^2$ on the right hand side. Then $\Pi_n(w : \|w - w_0\|_{L_2} > M\epsilon_n | X^{(n)}) \rightarrow 0$ in $P_{w_0}^{(n)}$ -probability, for some $M > 0$.*

5. The separable Hilbert space setting

In this section we consider p -exponential measures in a separable Hilbert space X . Since any separable Hilbert space is isometrically isomorphic to the space of square summable sequences ℓ_2 , we can equivalently, as far as concentration is concerned, work in ℓ_2 . This equivalence holds, provided the p -exponential measure in X is defined using expansions in an orthonormal basis, see Subsection 2.1. In particular, we consider α -regular p -exponential measures in sequence space and study their concentration at centers of varying Besov-type regularity. Note that these measures are merely a different parametrization of Besov-space priors used in applied inverse problems literature, [37]. We combine our findings with Theorem 4.2, to obtain posterior contraction rates in the white noise model under Besov-type regularity of the truth.

We first define the following Besov-type sequence spaces.

Definition 5.1. For any $s > 0$, $q \geq 1$ and $d \in \mathbb{N}$, we denote by B_q^s the weighted ℓ_q spaces

$$B_q^s = B_q^s(d) = \{u \in \mathbb{R}^\infty : \sum_{\ell=1}^{\infty} \ell^{q(\frac{s}{d} + \frac{1}{2})-1} |u_\ell|^q < \infty\}, \quad \|u\|_{B_q^s} = \left(\sum_{\ell=1}^{\infty} \ell^{q(\frac{s}{d} + \frac{1}{2})-1} |u_\ell|^q \right)^{\frac{1}{q}}.$$

The case $q = 2$ corresponds to Sobolev-type spaces. These spaces can be identified for example to Sobolev spaces H^s of periodic functions on $\mathbb{T}^d = (0, 1]^d$ with s square integrable derivatives, using expansions in the Fourier basis. Similarly, for $q \neq 2$ B_q^s can be identified with the Besov space $B_{q_1 q_2}^s$ of periodic functions on \mathbb{T}^d , with integrability parameters $q_1 = q_2 = q$ and smoothness parameter s , using expansions in certain sufficiently regular orthonormal wavelet bases [47]. Of particular interest is the case $q < 2$, which includes classes of non-smooth and spatially inhomogeneous functions, see for example [21]. Such functions are useful in many scientific disciplines, for example in geophysics and medical imaging, as they can be smooth in one area and rough in another one. The rates we obtain below suggest that, when interested in reconstructing an unknown function of this type, it is beneficial to use a non-Gaussian p -exponential prior.

5.1. α -regular p -exponential priors in ℓ_2

Consider μ a p -exponential measure in sequence space with $\gamma_\ell = \ell^{-\frac{1}{2} - \frac{\alpha}{d}}$, $\alpha > 0$, $p \in [1, 2]$, $d \in \mathbb{N}$. The parameter d expresses the dimension inherent in the Hilbert space X and is fixed for a given model, for example $X = L_2(\mathbb{T}^d)$. As discussed in Section 2.1, since $\gamma \in \ell_2$ it holds $\mu(\ell_2) = 1$. Furthermore, by Proposition 2.10 the support of μ is ℓ_2 . We call such a measure a d -dimensional α -regular p -exponential measure in ℓ_2 .

The next result studies the Besov-type regularity of draws from μ and justifies the name α -regular.

Lemma 5.2. Assume μ is a d -dimensional α -regular p -exponential measure in ℓ_2 . Then for any $q \geq 1$, we have that $\mu(B_q^s) = 1$ for all $s < \alpha$, and $\mu(B_q^s) = 0$ for all $s \geq \alpha$.

We next study the concentration function $\varphi_w(\cdot)$ of μ , defined for centers $w \in \ell_2$; see Definition 2.12 where $X = \ell_2$. The following lemma identifies the space \mathcal{Z} in which we approximate the center $w \in \ell_2$ in the first term of φ_w , as well as the shift space \mathcal{Q} . It follows immediately from Proposition 2.7 and Definitions 2.8 and 5.1.

Lemma 5.3. Assume μ is a d -dimensional α -regular p -exponential measure in ℓ_2 . Then $\mathcal{Z} = \mathcal{Z}_\alpha := B_p^{\alpha + \frac{d}{p}}$ and $\mathcal{Q} = \mathcal{Q}_\alpha := B_2^{\alpha + \frac{d}{2}}$.

We next study the blow-up rate of the concentration function $\varphi_w(\epsilon)$ as $\epsilon \rightarrow 0$ in the present setting. In particular, we find upper bounds on the minimal solution ϵ_n of the inequality $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$ depending on the Besov-type regularity of w_0 .

Proposition 5.4. Assume that μ is a d -dimensional α -regular p -exponential measure in ℓ_2 and that $w_0 \in B_q^\beta$ for $\beta > 0 \vee (\frac{d}{q} - \frac{d}{2})$, $q \geq 1$, $p \in [1, 2]$. Then as $n \rightarrow \infty$ the rate $\epsilon_n \asymp r_n^{\alpha, \beta, p, q}$ satisfies the inequality $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$, for constants which depend on w_0 only through its B_q^β norm and for $r_n^{\alpha, \beta, p, q}$ given below:

i) For $q \geq 2$

$$r_n^{\alpha, \beta, p, q} := \begin{cases} n^{-\frac{\beta}{d+2\beta+p(\alpha-\beta)}}, & \text{if } \alpha \geq \beta, \\ n^{-\frac{\alpha}{d+2\alpha}}, & \text{if } \alpha < \beta. \end{cases}$$

ii) For $q < 2$ and $p \leq q$, letting $a = \sqrt{2\beta dp + \beta^2 p^2 + d^2(1 + 2p - 4p/q)}$

$$r_n^{\alpha, \beta, p, q} := \begin{cases} n^{-\frac{2\beta q + d(q-2)}{4d(q-1) + 4\beta q + 2pq(\alpha-\beta)}}, & \text{if } \alpha \geq \frac{\beta p - d + a}{2p}, \\ n^{-\frac{\alpha}{d+2\alpha}}, & \text{if } \alpha < \frac{\beta p - d + a}{2p}. \end{cases}$$

iii) For $q < 2$ and $p > q$, letting $a = \sqrt{\frac{2\beta dq(2q-p) + \beta^2 pq^2 + d^2(p+2q^2-4q)}{p}}$

$$r_n^{\alpha, \beta, p, q} := \begin{cases} n^{-\frac{2\beta q + d(q-2)}{2d(p+q-2) + 4\beta q + 2pq(\alpha-\beta)}}, & \text{if } \alpha \geq \frac{\beta q - d + a}{2q}, \\ n^{-\frac{\alpha}{d+2\alpha}}, & \text{if } \alpha < \frac{\beta q - d + a}{2q}. \end{cases}$$

The proof of the last proposition, builds on Lemmas 5.12 and 5.13 in Subsection 5.3 below, in which we estimate the two terms of the concentration function. Notice that the assumption $\beta > 0 \vee (\frac{d}{q} - \frac{d}{2})$ secures that $w_0 \in \ell_2$, see Lemma G.1 in the supplement below.

Consider a nonparametric inference problem in a separable Hilbert parameter space X , where the X -norm relates suitably to the statistically relevant norms for the model and there exist exponentially powerful tests for separating the truth from balls in X at a certain distance from it. In the assumed separable Hilbert setting, we can verify that for ϵ_n the rate in Proposition 5.4, the maximum appearing in the right hand side of the complexity bound (7) in our general contraction Theorem 3.1, is dominated by $n\epsilon_n^2$; see Lemma 5.14 in Subsection 5.3. Together with Proposition 5.4 and Theorem 3.1, this suggests that if we use as prior a d -dimensional α -regular p -exponential measure in ℓ_2 identified with a measure on X via a series expansion in an orthonormal basis of X , then $r_n^{\alpha, \beta, p, q}$ is an upper bound on the posterior contraction rate when the truth belongs to B_q^β . Here we identify Besov regularity in X with Besov regularity of the sequence of coefficients in the orthonormal basis. For example, in the white noise model combining the last two results with Theorem 4.2, we get immediately the next result for α -regular p -exponential priors in $L_2[0, 1]$ (here $d = 1$).

Theorem 5.5. Consider the white noise model of Subsection 4.2, and let $\Pi = \mu$ be an α -regular p -exponential prior in $L_2[0, 1]$, $\alpha > 0$, $p \in [1, 2]$. Assume $w_0 \in B_q^\beta$, $\beta > 0 \vee (\frac{1}{q} - \frac{1}{2})$, $q \geq 1$ and let $r_n^{\alpha, \beta, p, q}$ be defined as in Proposition 5.4. Then for M large enough, as $n \rightarrow \infty$

$$\Pi_n(w \in L_2[0, 1] : \|w - w_0\|_{L_2} \leq Mr_n^{\alpha, \beta, p, q} |X^{(n)}|) \rightarrow 1,$$

in $P_{w_0}^{(n)}$ -probability.

Notice that since in Proposition 5.4 all the constants depend on w_0 only through its B_q^β -norm, the rates of contraction in the above theorem hold uniformly for w_0 in B_q^β -balls.

The last result generalizes existing contraction results in the conjugate setting of the white noise model with Gaussian priors and under Sobolev-type regularity of the truth, $p = q = 2$; see [62, Theorem 5.1] and [5, Theorem 2.1], as well as [12, Theorem 2] which discusses the sharpness of the Gaussian contraction rates. Note that in our setting, unless $p = 2$, the p -exponential prior is non-conjugate to the Gaussian likelihood of the white noise model. However, explicit calculations are possible to get upper bounds on the rate of posterior contraction, see [13, Corollary 3], for Sobolev-type regularity of the truth $q = 2$, when $\alpha \leq \beta$. Our result agrees with the existing rates in both aforementioned special cases, but goes further and in particular studies rates of contraction under Besov-type smoothness, that is the more intricate case $q \neq 2$.

Minimax rates in L_2 -loss, for function estimation in the white noise model under Besov-regularity, have been studied in [21]; see also [28, Chapter 10] for density estimation again in L_2 -loss. The results there, show that for all $q \geq 1$ and for $\beta > 1/q$ or $\beta \geq 1$ when $q = 1$, the minimax rate is

$$m_n := n^{-\frac{\beta}{1+2\beta}}. \quad (14)$$

An interesting feature, is that for $q < 2$ linear estimators do not achieve the minimax rate, and instead only achieve the rate

$$l_n := n^{-\frac{\beta-\gamma/2}{1+2\beta-\gamma}}, \quad (15)$$

where $\gamma := \frac{2-q}{q} > 0$. For $q \geq 2$, linear estimators do achieve the minimax rate m_n . This change of behaviour is attributed to the fact that, for $q < 2$, functions in B_q^β are not in general spatially homogeneous, but instead can be irregular in some parts and smooth in other parts. As explained in [21], linear estimators cannot cope well with this inhomogeneity and either oversmooth the irregular part, or undersmooth the smooth part, or both.

Remark 5.6 (Results for $q \geq 2$). *An inspection of the bounds in Theorem 5.5, reveals that for $q \geq 2$, the particular value of q does not influence the contraction rate. When $\beta = \alpha$, we get the minimax rate m_n , see (14), independently of $p \in [1, 2]$. In the case of an undersmoothing prior, $\beta > \alpha$, the rates for all $p \in [1, 2]$ coincide and are slower than the minimax rate. Finally, for an oversmoothing prior, $\beta < \alpha$, the rate is faster the smaller p is. This is reasonable, since for smaller p there is a higher probability of ξ_ℓ in the definition of the p -exponential prior having large values, which counteracts the oversmoothing effect of the prior-scaling sequence $\gamma_\ell = \ell^{-\frac{1}{2}-\alpha}$.*

Remark 5.7 (Results for $q < 2$). *An inspection of the bounds in Theorem 5.5, reveals that for $q < 2$, the particular value of q does influence the contraction rate and we do not get the minimax rate m_n defined in (14), for any admissible*

combination of α, β, p, q . The best rates are achieved for $p = q$ and are better than the linear minimax l_n , see (15), while Gaussian priors appear to have the worst performance and to be limited by the linear minimax rate:

- i) Let $q < 2$, $p \leq q$ and fix $\beta > \frac{1}{q}$ or $\beta \geq 1$ if $q = 1$ (so that the minimax and linear minimax rates in [21] hold). For small $\alpha > 0$, the rate is $n^{-\frac{\alpha}{1+2\alpha}}$ which is suboptimal but improves as α increases. Since the other leg of the bound deteriorates as α increases, to achieve the minimax rate m_n , the bound $n^{-\frac{\alpha}{1+2\alpha}}$ needs to hold for α all the way up to $\alpha = \beta$. An easy calculation shows that the switching point between the two legs, $\alpha = \frac{\beta p - 1 + a}{2p}$, is smaller than β but larger than $\beta - \frac{\gamma}{2}$. This implies that we cannot achieve the minimax rate, but there are values of α for which we achieve rates faster than the linear minimax rate l_n . In fact, one can check that for larger p (that is $p \leq q$ closer to $q < 2$), the switching point gets closer to β , hence we can get closer to the minimax rate for suitable values of α . However, even for $p = q$ we cannot reach the minimax rate.
- ii) Let $q < 2$, $p > q$ and fix $\beta > \frac{1}{q}$ or $\beta \geq 1$ if $q = 1$. The reasoning is the same as in item (i): the switching point in the rate, $\frac{\beta q - 1 + a}{2q}$, is for all $p \in (q, 2]$ smaller than β , for $p \in (q, 2)$ larger than $\beta - \frac{\gamma}{2}$ and for $p = 2$ equal to $\beta - \frac{\gamma}{2}$. Hence we cannot achieve the minimax rate m_n for any $p \in (q, 2]$, for $p \in (q, 2)$ there are values of α achieving rates faster than the linear minimax rate l_n and for the Gaussian case, $p = 2$, the best we can achieve is the linear minimax rate l_n . One can check that for p smaller (that is p closer to q), the switching point gets closer to β , hence we can get closer to the minimax rate for suitable values of α . Note, that at this stage it is not clear whether Gaussian priors are fundamentally limited by the linear minimax rate, see the discussion in Section H of the supplement below.

As detailed in the last remark, for $w_0 \in B_q^\beta$ with $q < 2$, for all $p \in [1, 2]$, the best rate is achieved by an undersmoothing prior and this rate is not minimax. This is due to the infimum term in the concentration function dominating the centered small ball probability term, already for large enough $\alpha < \beta$. This motivates using *rescaled* p -exponential priors in the next subsection, with a vanishing scaling as $n \rightarrow \infty$: the idea is to use an undersmoothing prior such that without rescaling the centered small ball probability term dominates in the concentration function, and choose the scaling to - in some sense - infuse additional regularity to the prior, in particular in order to balance the two terms of the concentration function. We will see that this can lead to minimax rates.

5.2. Rescaled α -regular p -exponential priors in ℓ_2

Consider $\bar{\mu}$ to be a rescaling of a d -dimensional α -regular p -exponential prior μ in ℓ_2 , that is we take $\bar{\gamma}_\ell = \lambda \gamma_\ell$ where $\gamma_\ell = \ell^{-\frac{1}{2} - \frac{\alpha}{d}}$, for some $\lambda > 0$. Then $\mathcal{Z} = \bar{\mathcal{Z}}_\alpha$, $\mathcal{Q} = \bar{\mathcal{Q}}_\alpha$, where $\bar{\mathcal{Z}}_\alpha, \bar{\mathcal{Q}}_\alpha$ coincide with the spaces $\mathcal{Z}_\alpha, \mathcal{Q}_\alpha$ in Lemma 5.3,

but with norms $\|\cdot\|_{\tilde{\mathcal{Z}}_\alpha} = \lambda^{-1} \|\cdot\|_{\mathcal{Z}_\alpha}$ and $\|\cdot\|_{\tilde{\mathcal{Q}}_\alpha} = \lambda^{-1} \|\cdot\|_{\mathcal{Q}_\alpha}$, respectively. For any $w \in X$, the concentration function of $\bar{\mu}$ is then given by

$$\bar{\varphi}_w(\epsilon) = \lambda^{-p} \inf_{h \in \mathcal{Z}_\alpha: \|h-w\|_{\ell_2} \leq \epsilon} \|h\|_{\mathcal{Z}_\alpha}^p - \log \mu \left(\frac{\epsilon}{\lambda} B_{\ell_2} \right). \quad (16)$$

In the formulation of our results below, we let as in the discussion in the previous subsection

$$m_n := n^{-\frac{\beta}{d+2\beta}}, \quad (17)$$

which is the minimax rate of estimation of a $B_q^\beta(d)$ sequence under Gaussian white noise in ℓ_2 -loss, for all $q \geq 1$ and for $\beta > \frac{d}{q}$ or $\beta \geq d$ when $q = 1$, see [21].

We focus on the case $q < 2$, since for $q \geq 2$ we can achieve the minimax rate with α -regular p -exponential priors, for $\alpha = \beta$ without rescaling, see Remark 5.6. For $q \geq 2$, rescaling can still be beneficial, in particular in order to achieve (some degree of) adaptation to unknown smoothness β ; for results for Gaussian priors see [34, 53, 46]. We will investigate adaptation with p -exponential priors in a separate study.

Proposition 5.8. *Assume that $\bar{\mu} = \bar{\mu}_n$ is a rescaled d -dimensional α -regular p -exponential measure in ℓ_2 , $p \in [1, 2]$, corresponding to $\lambda = \lambda_n$. Let $w_0 \in B_q^\beta$ for $\beta > \frac{d}{p} \vee \frac{d}{q}$, $q \in [1, 2]$, and consider the corresponding minimax rate of estimation in Gaussian white noise in ℓ_2 -loss, m_n , as defined in (17). Then as $n \rightarrow \infty$, the rate $\epsilon_n \asymp \bar{r}_n$ satisfies the inequality $\bar{\varphi}_{w_0}(\epsilon_n) \leq n\epsilon_n^2$, for constants which depend on w_0 only through its B_q^β norm, where $\bar{r}_n = \bar{r}_n^{\alpha, \beta, p, q}$ and $\lambda_n = \lambda_n^{\alpha, \beta, p, q}$ are given below:*

- i) If $q = p$, $\alpha = \beta - \frac{d}{p}$, then $\bar{r}_n = m_n$ for $\lambda_n = n^{-\frac{d}{p(d+2\beta)}}$.
- ii) If $q > p$, $\alpha = \beta - \frac{d}{p}$, then $\bar{r}_n = m_n \log^{\frac{d(q-p)}{pq(d+2\beta)}} n$ for $\lambda_n = n^{-\frac{d}{p(d+2\beta)}} \log^\omega n$, where $\omega = (p - \frac{2d}{d+2\beta}) \frac{q-p}{p^2q} > 0$.

In all other combinations of α, β, p, q , for any choice of λ_n , \bar{r}_n is polynomially slower than m_n . In particular, if $q < p$, then the best achievable rate is $\bar{r}_n = n^{\frac{d(p-q)-\beta pq}{2d(q-p)+2\beta pq+pqd}}$, for $\alpha = \beta - \frac{d}{q}$ and $\lambda_n = n^{-\frac{qd}{2qd+2\beta pq-2pd+pqd}}$.

Notice, that the assumption $\beta > \frac{d}{p} \vee \frac{d}{q}$, is in place in order to secure that, for a fixed β , the values of α corresponding to the best achievable rates, are positive hence admissible, simultaneously for all combinations of p, q . This facilitates the comparison between the different choices of p, q . However, it is possible to compute (slower than minimax) rates with rescaled p -exponential priors for smaller values of $\beta > 0 \vee (\frac{d}{q} - \frac{d}{2})$, for $\alpha > 0$, with the same techniques used in the proof of Proposition 5.8.

As in the previous subsection, we can verify that for ϵ_n the rate in Proposition 5.8, the maximum appearing in the right hand side of the complexity bound (7) in our general contraction Theorem 3.1, is dominated by $n\epsilon_n^2$; see Lemma 5.15 in Subsection 5.3. This can be combined with Proposition 5.8 and Theorem 3.1, in order to get contraction rates in suitable nonparametric problems in separable

Hilbert spaces, under rescaled d -dimensional p -exponential priors. In particular, in the white noise model, combining the last two results with Theorem 4.2, we get immediately the next result for rescaled α -regular p -exponential priors in $L_2[0, 1]$ (here $d = 1$).

Theorem 5.9. *Consider the white noise model of Subsection 4.2, and let $\Pi_n = \bar{\mu}_n$ be a rescaled α -regular p -exponential prior in $L_2[0, 1]$, $\alpha > 0, p \in [1, 2]$. Assume $w_0 \in B_q^\beta$, $\beta > \frac{d}{p} \vee \frac{d}{q}$, $q \in [1, 2)$ and let \bar{r}_n and λ_n be defined as in Proposition 5.8. Then for M large enough, as $n \rightarrow \infty$*

$$\Pi_n(w \in L_2[0, 1] : \|w - w_0\|_{L_2} \leq M\bar{r}_n |X^{(n)}|) \rightarrow 1,$$

in $P_{w_0}^{(n)}$ -probability.

Remark 5.10. *The results of Theorem 5.9 show again that for $w_0 \in B_q^\beta$ with $q \in [1, 2)$, $\beta > \frac{1}{p} \vee \frac{1}{q}$, the best rates are achieved for $p = q$, while Gaussian priors appear to have the worst performance and to be limited by the linear minimax rate:*

- i) *We can achieve the minimax rate m_n , see (14), only with a p -exponential prior for $p = q$, for regularity $\alpha = \beta - \frac{1}{p}$ and for appropriate vanishing rescaling λ_n . For $p < q$ we can achieve the minimax rate m_n up to logarithmic factors, again for regularity $\alpha = \beta - \frac{1}{p}$ and an appropriate vanishing rescaling.*
- ii) *For $q < p < 2$, the best rate with rescaling is obtained for $\alpha = \beta - \frac{1}{q}$ and is improved compared to the best obtained rate without rescaling, but it remains polynomially slower than the minimax rate m_n , defined in (14). The fact that the rate improves, is implicit in the proof of Proposition 5.8: it is shown that for any $\alpha \neq \beta - \frac{1}{q}$ the obtained rate with optimized rescaling is strictly slower than for $\alpha = \beta - \frac{1}{q}$, while as discussed in Remark 5.7(ii), without rescaling the best rate is achieved for $\alpha > \beta - \frac{\gamma}{2} = \beta - \frac{1}{q} + \frac{1}{2}$, in particular for $\alpha \neq \beta - \frac{1}{q}$.*
- iii) *For the Gaussian case, $p = 2$, rescaling does not improve the best obtainable rate, which is the linear minimax rate l_n defined in (15). As can be seen at the end of the proof of Proposition 5.8, this rate is achieved for any $\alpha \geq \beta - \frac{1}{q}$ for appropriate rescaling (or for no rescaling, when $\alpha = \beta - \frac{\gamma}{2}$, in agreement with Remark 5.7(ii)). We reiterate, that at this stage it is not clear whether Gaussian priors are fundamentally limited by the linear minimax rate, see the discussion in Section H of the supplementary material below.*
- iv) *The case $w_0 \in B_1^1$, which is particularly interesting in applications like signal processing, is not immediately covered by Theorem 5.9. However, an inspection of the proof of Proposition 5.8 shows that using a rescaled α -regular p -exponential prior, with $p = 1$ and for $\alpha > 0$ arbitrarily small, we can achieve a rate of contraction arbitrarily close to the minimax rate m_n in (14).*

Remark 5.11. The product Laplace prior ($p = 1$ and $\gamma_\ell = 1, \forall \ell \in \mathbb{N}$), has been studied in the sparse Gaussian sequence model in [14, Section 3]. Even though the posterior mode corresponds to the LASSO, which is known to provide minimax optimal estimation in this setting, [14, Theorem 7] shows that the whole posterior contracts at a suboptimal rate for truly sparse signals. This is because the posterior variance is overly large. On the other hand, Theorem 5.9 shows that in the white noise model, under sparsity assumptions expressed in terms of Besov regularity with integrability parameter $q < 2$, the appropriately tuned rescaled α -regular Laplace prior results in a posterior which contracts optimally for $q = 1$ or optimally up to logarithmic terms for $q \neq 1$.

5.3. Estimates relating to the concentration function

5.3.1. α -regular p -exponential priors in ℓ_2

We first study the centered small ball probability term in the concentration function. The result is a direct consequence of [4, Theorem 4.2].

Lemma 5.12. Assume μ is a d -dimensional α -regular p -exponential measure in ℓ_2 . Then as $\epsilon \rightarrow 0$

$$-\log \mu(\epsilon B_{\ell_2}) \asymp \epsilon^{-\frac{d}{\alpha}}.$$

In the next lemma we compute upper bounds on the first term in the concentration function φ_w , depending on the Besov regularity of a $w \in \ell_2$.

Lemma 5.13. Assume that μ is a d -dimensional α -regular p -exponential measure in ℓ_2 and that $w_0 \in B_q^\beta$ for $\beta > 0 \vee (\frac{d}{q} - \frac{d}{2})$, $q \geq 1$. Then as $\epsilon \rightarrow 0$, we have the following bounds where all the constants depend on w_0 only through its B_q^β -norm:

i) For $q \leq p$ (hence $q \leq 2$)

$$\inf_{h \in \mathcal{Z}_\alpha: \|h - w_0\|_{\ell_2} \leq \epsilon} \|h\|_{\mathcal{Z}_\alpha}^p \lesssim \begin{cases} 1, & \text{if } \beta \geq \alpha + \frac{d}{q}, \\ \epsilon^{2p \frac{(\beta - \alpha)q - d}{(2\beta + d)q - 2d}}, & \text{if } \beta < \alpha + \frac{d}{q}. \end{cases}$$

ii) For $q > p$

$$\inf_{h \in \mathcal{Z}_\alpha: \|h - w_0\|_{\ell_2} \leq \epsilon} \|h\|_{\mathcal{Z}_\alpha}^p \lesssim \begin{cases} 1, & \text{if } \beta > \alpha + \frac{d}{p}, \\ (-\log \epsilon)^{\frac{q-p}{q}}, & \text{if } \beta = \alpha + \frac{d}{p}, \\ \epsilon^{\frac{\beta p - \alpha p - d}{\beta}}, & \text{if } \beta < \alpha + \frac{d}{p} \text{ and } q \geq 2, \\ \epsilon^{2q \frac{(\beta - \alpha)p - d}{(2\beta + d)q - 2d}}, & \text{if } \beta < \alpha + \frac{d}{p} \text{ and } q < 2. \end{cases}$$

We finally verify that for ϵ_n the rate in Proposition 5.4, the quality of approximation of $\mathcal{Q} = \mathcal{Q}_\alpha$ by $\mathcal{Z} = \mathcal{Z}_\alpha$ in ℓ_2 , is sufficiently good for the maximum appearing in the right hand side of the complexity bound (7) in our general contraction Theorem 3.1, to be dominated by $n\epsilon_n^2$.

Lemma 5.14. *Let μ be a d -dimensional α -regular p -exponential measure in ℓ_2 . For $\epsilon, a > 0$, define*

$$f(a) = a^p(1 \vee a^{\frac{2d-pd}{d+2\alpha}}) \text{ and } g(\epsilon) = 2(1 \vee \epsilon^{-\frac{2d}{d+2\alpha}}). \quad (18)$$

Then f and g satisfy the approximation bound (6) in Section 3 and, moreover,

$$f(n^{\frac{1}{2}}\epsilon_n)g(\epsilon_n)^{1-\frac{p}{2}} \lesssim n\epsilon_n^2 \quad (19)$$

for all $p \in [1, 2]$, where $\epsilon_n = r_n^{\alpha, \beta, p, q}$ for $r_n^{\alpha, \beta, p, q}$ as in Proposition 5.4.

5.3.2. Rescaled α -regular p -exponential priors in ℓ_2

The expression (16) for the concentration function $\bar{\varphi}_w(\cdot)$ of rescaled α -regular p -exponential priors, suggests that we can re-use Lemmas 5.12 and 5.13 in order to prove Proposition 5.8. In the next lemma we also verify that for ϵ_n the rate in Proposition 5.8, the quality of approximation of $\mathcal{Q} = \bar{\mathcal{Q}}_\alpha$ by $\mathcal{Z} = \bar{\mathcal{Z}}_\alpha$ in ℓ_2 , is sufficiently good for the maximum appearing in the right hand side of the complexity bound (7) in our general contraction Theorem 3.1, to be dominated by $n\epsilon_n^2$.

Lemma 5.15. *Let $\bar{\mu}$ be a rescaled d -dimensional α -regular p -exponential measure in ℓ_2 . For $t > 0$, define $\bar{f}(t) := \lambda^{-p}f(\lambda t)$ and $\bar{g}(t) := g(t)$, where f, g are as in Lemma 5.14. Then \bar{f} and \bar{g} satisfy the approximation bound (6) in Section 3, and, moreover,*

$$\bar{f}(n^{\frac{1}{2}}\epsilon_n)\bar{g}(\epsilon_n)^{1-\frac{p}{2}} \lesssim n\epsilon_n^2 \quad (20)$$

for all $p \in [1, 2]$, where $\epsilon_n = \bar{r}_n$ and $\lambda = \lambda_n$ as in Proposition 5.8.

6. The $C[0, 1]$ setting

In this section we consider p -exponential measures in the separable Banach space $X = C[0, 1]$. We define p -exponential measures using an appropriately regular Schauder basis, see below for details. In particular, we define α -regular p -exponential measures in $C[0, 1]$ and study their concentration at centers of varying Hölder-type regularity. We combine our findings with Theorem 4.1, to obtain posterior contraction rates for density estimation under Hölder-type regularity of the truth.

We consider orthonormal wavelet bases of $L_2[0, 1]$, constructed as discussed in [15]; see [27] or [41] for fundamentals of wavelet analysis. We denote such a wavelet basis by $\{\psi_{kl} : k \in \mathbb{N}_0, l = 1, \dots, 2^k\}$, where k corresponds to the resolution level and l to the location. A function $u \in L_2[0, 1]$ can be expanded as $\sum_{k=0}^{\infty} \sum_{l=1}^{2^k} u_{kl} \psi_{kl}$, where the coefficients u_{kl} are given by the L_2 -inner products between u and ψ_{kl} . We assume that ψ_{kl} are S -Hölder continuous for some $S > 0$. We record some properties that will be useful for our analysis, see [15, 27]:

- $\{\psi_{kl}\}$ is a Schauder basis of $C[0, 1]$.

- There exists a constant $C_1 > 0$ such that

$$|\psi_{kl}(x) - \psi_{kl}(y)| \leq C_1 2^{\frac{k}{2} + k\vartheta} |x - y|^\vartheta, \quad \vartheta \leq S \wedge 1. \quad (21)$$

- There exists a constant $C_2 > 0$ such that

$$\left\| \sum_{l=1}^{2^k} u_{kl} \psi_{kl} \right\|_{L_\infty} \leq C_2 2^{\frac{k}{2}} \sup_{1 \leq l \leq 2^k} |u_{kl}|. \quad (22)$$

- Let $0 < s < S$. Then u belongs to the Besov space $B_{\infty\infty}^s[0, 1]$ if and only if

$$\|u\|_{B_{\infty\infty}^s} := \sup_{k \geq 0; 1 \leq l \leq 2^k} 2^{k(\frac{1}{2} + s)} |u_{kl}| < \infty.$$

Furthermore, if s is non-integer we have that $g \in C^s$ if and only if

$$\|u\|_{B_{\infty\infty}^s} < \infty.$$

Note, that our analysis holds for other possibly nonorthonormal multiresolution Schauder bases, provided the above bounds on ψ_{kl} and the characterizations in terms of the coefficients u_{kl} hold. For example, one can use the Faber (integrated Haar) basis, see [56, Section 3.1.3]. We use basis functions ψ_{kl} which have sufficient Hölder regularity, so that ψ_{kl} can characterize the maximal (s, ∞, ∞) -Besov (or s -Hölder) regularity we consider, that is we assume $S > \max\{\alpha, \beta\}$, where α, β will express the regularity of the prior and truth, respectively.

We can define a p -exponential measure μ in $C[0, 1]$ by randomizing the coefficients in the expansion

$$u(t) = \sum_{k=0}^{\infty} \sum_{l=1}^{2^k} u_{kl} \psi_{kl}(t), \quad t \in [0, 1].$$

We let

$$u_{kl} = \gamma_{kl} \xi_{kl}, \quad \xi_{kl} \stackrel{iid}{\sim} f_p, \quad p \in [1, 2], \quad \gamma_{kl} = 2^{-(\frac{1}{2} + \alpha)k}, \quad \alpha > 0. \quad (23)$$

The next result studies Hölder continuity of draws from this p -exponential measure.

Proposition 6.1. *Let μ be the p -exponential measure defined in (23), for any $p \in [1, 2]$ and $\alpha > 0$. Then $\mu(C^s) = 1$ for all $s < \alpha \wedge 1$.¹*

In particular, the last proposition implies that indeed μ is a measure on $X = C[0, 1]$. We call μ defined in (23) an α -regular p -exponential measure in $C[0, 1]$.

¹By assuming additional regularity on the basis functions and using techniques relying on the embeddings of Besov spaces, it is possible to remove the requirement $s < 1$ in Proposition 6.1. See [17, Theorem 7] for a similar derivation and [55, Section 4.6.1] for the relevant embeddings.

By Proposition 2.10 the topological support of μ is the space $C[0, 1]$. We next study the concentration function $\varphi_w(\cdot)$ of μ , defined for centers $w \in C[0, 1]$; see Definition 2.12 where $X = C[0, 1]$, with $\|\cdot\|_X = \|\cdot\|_{L_\infty}$. The next lemma identifies the space \mathcal{Z} in which we approximate the center $w \in C[0, 1]$ in the first term of φ_w , as well as the shift space \mathcal{Q} . Note that these spaces can be defined in sequence space, independently of the parameter space X and the Schauder basis in which we work. The lemma follows immediately by Proposition 2.7 and Definition 2.8.

Lemma 6.2. *Assume μ is an α -regular p -exponential measure in $C[0, 1]$. Then*

$$\mathcal{Z} = \mathcal{Z}_\alpha := \{h \in \mathbb{R}^\infty : \sum_{k=1}^{\infty} \sum_{l=1}^{2^k} |h_{kl}|^{p 2^{(\frac{1}{2}+\alpha)pk}} < \infty\}, \quad \|h\|_{\mathcal{Z}_\alpha} = \left(\sum_{k=1}^{\infty} \sum_{l=1}^{2^k} |h_{kl}|^{p 2^{(\frac{1}{2}+\alpha)pk}} \right)^{\frac{1}{p}},$$

$$\mathcal{Q} = \mathcal{Q}_\alpha := \{h \in \mathbb{R}^\infty : \sum_{k=1}^{\infty} \sum_{l=1}^{2^k} h_{kl}^2 2^{(1+2\alpha)k} < \infty\}, \quad \|h\|_{\mathcal{Q}_\alpha} = \left(\sum_{k=1}^{\infty} \sum_{l=1}^{2^k} h_{kl}^2 2^{(1+2\alpha)k} \right)^{\frac{1}{2}}.$$

In fact, due to the asymptotic equivalence of the sequences $\gamma_{kl} = 2^{-(\frac{1}{2}+\alpha)k}$, $k \in \mathbb{N}, 1 \leq l \leq 2^k$ and $\gamma_\ell = \ell^{-\frac{1}{2}-\alpha}, \ell \in \mathbb{N}$, we have that $\mathcal{Z}_\alpha = B_p^{\alpha+\frac{1}{p}}$ and $\mathcal{Q}_\alpha = B_2^{\alpha+\frac{1}{2}}$, where B_q^s are the Besov-type spaces of sequences defined in Definition 5.1 for $d = 1$.

We next study the centered small ball probability term in the concentration function. For the proof we use the techniques of [51] which studies the Gaussian case.

Proposition 6.3. *Let μ be an α -regular p -exponential measure in $C[0, 1]$. Then as $\epsilon \rightarrow 0$*

$$-\log \mu(\{u \in C[0, 1] : \|u\|_{L_\infty} \leq \epsilon\}) \lesssim \epsilon^{-\frac{1}{\alpha}}.$$

Finally, in the next lemma we compute upper bounds on the first term in the concentration function φ_w , depending on the (β, ∞, ∞) -Besov regularity of w , which recall is identified with β -Hölder regularity when β is non-integer.

Lemma 6.4. *Assume that μ is an α -regular p -exponential measure in $C[0, 1]$ and that $w_0 \in B_{\infty\infty}^\beta$, $\beta > 0$. Then, as $\epsilon \rightarrow 0$*

$$\inf_{h \in \mathcal{Z}_\alpha : \|h - w_0\|_{L_\infty} \leq \epsilon} \|h\|_{\mathcal{Z}_\alpha}^p \lesssim \begin{cases} \epsilon^{\frac{\beta p - \alpha p - 1}{\beta}}, & \text{if } \beta < \alpha + \frac{1}{p}, \\ \log(1/\epsilon), & \text{if } \beta = \alpha + \frac{1}{p}, \\ 1, & \text{if } \beta > \alpha + \frac{1}{p}. \end{cases}$$

Combining the previous lemmas, we can find upper bounds on the minimal solution ϵ_n of the inequality $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$ depending on the Hölder regularity of w_0 . Since the rates on the right hand sides of the bounds in Proposition 6.3 and Lemma 6.4 are identical to the ones in Lemmas 5.12 and 5.13 (for $d = 1$ and $q \geq 2$), respectively, the proof is identical to the proof of part (i) of Proposition 5.4 and is hence omitted.

Proposition 6.5. Assume that μ is an α -regular p -exponential measure in $C[0, 1]$ and that $w_0 \in B_{\infty\infty}^\beta$, $\beta > 0$. Then as $n \rightarrow \infty$ the rate $\epsilon_n \asymp \rho_n^{\alpha,\beta,p}$ satisfies the inequality $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$, where

$$\rho_n^{\alpha,\beta,p} := \begin{cases} n^{-\frac{\beta}{1+2\beta+p(\alpha-\beta)}}, & \text{if } \beta \leq \alpha, \\ n^{-\frac{\alpha}{1+2\alpha}}, & \text{if } \beta > \alpha. \end{cases}$$

We next study the quality of the approximation of elements of \mathcal{Q}_α by elements of \mathcal{Z}_α in the supremum norm, that is, determine functions g, f such that (6) in Section 3 holds.

Lemma 6.6. Let μ be an α -regular p -exponential measure in $C[0, 1]$. Then there exists $c > 0$ depending only on the Schauder basis, p and α , such that the functions $f(a) = ca^{\frac{2-p+2\alpha p}{2\alpha}}$ and $g(\epsilon) = \epsilon^{-\frac{1}{\alpha}}$ satisfy (6).

A straightforward computation shows that for the above g and f , the rate $\epsilon_n = \rho_n^{\alpha,\beta,p}$ is such that the right hand side of the complexity bound (7) in Theorem 3.1 is dominated by $n\epsilon_n^2$ only if $\beta \leq \alpha - \frac{1}{2-p+2\alpha p}$. This means that the complexity bound we obtain from Theorem 3.1 does not match the conditions of general results like [25, Theorems 8.9 and 8.19] and we need to use Corollary 3.2 to get contraction rates. To this end we solve (10) and find that for these functions f, g and for $\epsilon_n = \rho_n^{\alpha,\beta,p}$, the fastest decaying solution is $\tilde{\epsilon}_n \asymp \tilde{\rho}_n^{\alpha,\beta,p}$, where

$$\tilde{\rho}_n^{\alpha,\beta,p} := \begin{cases} n^{\frac{(2-p)(1-2\alpha)}{8\alpha} - \frac{p\beta}{2(1+2\beta+p(\alpha-\beta))}}, & \text{if } \beta \leq \alpha, \\ n^{\frac{2-p-8\alpha^2}{8\alpha(1+2\alpha)}}, & \text{if } \beta > \alpha. \end{cases} \quad (24)$$

For a fixed value of the regularity of the truth $\beta > 0$, note that $\tilde{\rho}_n^{\alpha,\beta,p}$ decays only for sufficiently large prior regularity α . For example, if $\alpha < \beta$, we have decay only for $\alpha > \sqrt{\frac{2-p}{8}}$. As $p \rightarrow 2$, since the difficulty in the complexity bound (7) disappears, the rates $\tilde{\rho}_n^{\alpha,\beta,p}$ approach the rates $\rho_n^{\alpha,\beta,p}$.

For example, combining these considerations with Theorem 4.1, we get immediately the following result giving contraction rates for density estimation.

Theorem 6.7. Consider the density estimation model of Subsection 4.1, and let W be an α -regular p -exponential random element in $C[0, 1]$, $\alpha > 0, p \in [1, 2]$. Assume $w_0 = \log \pi_0 \in B_{\infty\infty}^\beta$, $\beta > 0$ and denote by P_0^n the distribution of the vector (X_1, \dots, X_n) . Let $\rho_n^{\alpha,\beta,p}, \tilde{\rho}_n^{\alpha,\beta,p}$ be defined as in Proposition 6.5 and (24), respectively. Then for M large enough, as $n \rightarrow \infty$

$$\Pi_n(\pi : d_H(\pi, \pi_0) > M(\rho_n^{\alpha,\beta,p} \vee \tilde{\rho}_n^{\alpha,\beta,p}) | X_1, \dots, X_n) \rightarrow 0,$$

in P_0^n -probability.

For $p = 2$, we have that $\rho_n^{\alpha,\beta,2} = \tilde{\rho}_n^{\alpha,\beta,2}$ and we recover existing contraction rates for Gaussian priors, see [25, Section 11.4] or [27, Theorem 7.3.9]. In this case, if the regularity of the prior matches the regularity of the truth $\alpha = \beta$, we get the minimax estimation rate in the Hellinger distance for functions which

are β -Hölder continuous, $n^{-\frac{\beta}{1+2\beta}}$. For $p \in [1, 2)$, a straightforward calculation shows that the rate $\tilde{\rho}_n^{\alpha, \beta, p}$ is slower than $\rho_n^{\alpha, \beta, p}$ unless $0 < \beta \leq \alpha - \frac{1}{2-p+2\alpha p}$. In particular, for $\alpha = \beta$ we only have contraction if the prior is sufficiently regular, $\alpha > \sqrt{\frac{2-p}{8}}$, with contraction rate $\tilde{\rho}_n^{\alpha, \alpha, p}$ which is slower than the minimax rate. As p increases towards $p = 2$ the gap disappears; likewise for large α .

It appears that contrary to the Gaussian case $p = 2$, studying prior concentration and using general contraction results relying on prior mass and entropy conditions, is not optimal for proving contraction rates for p -exponential priors in $C[0, 1]$ when $p \in [1, 2)$. This is due to the more complicated complexity bound (7) compared to the Gaussian case, which in this setting affects the rates because of the poor approximation quality of \mathcal{Q}_α by \mathcal{Z}_α in the supremum norm. Note that rescaling the prior as considered in Subsection 5.2, does not help with this issue. In general contraction results like [25, Theorem 8.9], the entropy condition is used to construct certain necessary tests. Directly constructing the necessary tests and using other general contraction results which do not rely on entropy conditions, for example [25, Theorem 8.12], may resolve this issue. This is out of the scope of the present paper, but it is a possible future direction.

Acknowledgements The authors would like to thank Emanuel Milman and Mikhail Lifshits for providing very useful guidance regarding isoperimetry and small ball probabilities, respectively. SA would also like to thank Richard Nickl, Ismael Castillo, Kolyan Ray and Aad van der Vaart for answering many questions and providing important insight regarding posterior contraction rates. SA is particularly grateful to Sven Wang for sharing the concept of using rescaling for improving rates rather than just for adaptation, and to Richard Nickl for providing a very stimulating research environment during a research visit to Cambridge. Finally, the authors are thankful to an anonymous associate editor and two anonymous referees, for corrections and suggestions that led to a substantial improvement of some of the results in this article.

References

- [1] S. Agapiou, M. Burger, M. Dashti, and T. Helin. Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems. *Inverse Problems*, 34(4):045002, 2018.
- [2] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart. Importance sampling: Intrinsic dimension and computational cost. *Statistical Science*, 32(3):405–431, 2017.
- [3] J. Arbel, G. Gayraud, and J. Rousseau. Bayesian optimal adaptive estimation using a sieve prior. *Scandinavian journal of statistics*, 40(3):549–570, 2013.
- [4] F. Aurzada. On the lower tail probabilities of some random sequences in ℓ_p . *Journal of Theoretical Probability*, 20(4):843–858, 2007.
- [5] E. Belitser and S. Ghosal. Adaptive bayesian inference on the mean of an infinite-dimensional normal distribution. *The Annals of Statistics*, 31(2):536–559, 2003.

- [6] V. I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.
- [7] V. I. Bogachev. *Differentiable measures and the Malliavin calculus*, volume 164 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2010.
- [8] C. Borell. Convex measures on locally convex spaces. *Arkiv för Matematik*, 12(1):239–252, 1974.
- [9] C. Borell. The brunn-minkowski inequality in gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975.
- [10] E. Candes and J. Romberg. l1-magic: Recovery of sparse signals via convex programming. URL: www.acm.caltech.edu/l1magic/downloads/l1magic.pdf, 4:14, 2005.
- [11] E. J. Candes and D. L. Donoho. Curvelets: A surprisingly effective non-adaptive representation for objects with edges. Technical report, Stanford University, Department of Statistics, 2000.
- [12] I. Castillo. Lower bounds for posterior rates with gaussian process priors. *Electronic Journal of Statistics*, 2:1281–1299, 2008.
- [13] I. Castillo and R. Nickl. Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *The Annals of Statistics*, 41(4):1999–2028, 2013.
- [14] I. Castillo, J. Schmidt-Hieber, and A. van der Vaart. Bayesian linear regression with sparse priors. *Ann. Statist.*, 43(5):1986–2018, 2015.
- [15] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Applied and computational harmonic analysis*, 1993.
- [16] G. Da Prato. *An introduction to infinite-dimensional analysis*. Springer Science & Business Media, 2006.
- [17] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In *Handbook of uncertainty quantification. Vol. 1, 2, 3*, pages 311–428. Springer, Cham, 2017.
- [18] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.
- [19] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.
- [20] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.
- [21] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26(3):879–921, 1998.
- [22] D. L. Donoho, R. C. Liu, and B. MacGibbon. Minimax risk over hyperrectangles, and implications. *Ann. Statist.*, 18(3):1416–1437, 1990.
- [23] S. Ghosal, J. K. Ghosh, and A. W. Van Der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, pages 500–531, 2000.
- [24] S. Ghosal and A. Van Der Vaart. Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, 35(1):192–223, 2007.

- [25] S. Ghosal and A. van der Vaart. *Fundamentals of nonparametric Bayesian inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.
- [26] E. Giné and R. Nickl. Rates of contraction for posterior distributions in l_r -metrics, 1 r. *The Annals of Statistics*, 39(6):2883–2911, 2011.
- [27] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press, 2016.
- [28] W. Härdle, G. Kerkycharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.
- [29] T. Helin and M. Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8):085009, 2015.
- [30] I. M. Johnstone. Minimax bayes, asymptotic minimax and sparse wavelet priors. In *Statistical Decision Theory and Related Topics V*, pages 303–326. Springer, 1994.
- [31] A. Kagan. Statistical approach to some mathematical problems. *Austrian Journal of Statistics*, 32:71–83, 2003.
- [32] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006.
- [33] G. S. Kimeldorf and G. Wahba. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2):495–502, 1970.
- [34] B. T. Knapik, B. T. Szabó, A. W. van der Vaart, and J. H. van Zanten. Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probab. Theory Related Fields*, 164(3-4):771–813, 2016.
- [35] B. T. Knapik, A. W. van Der Vaart, and J. H. van Zanten. Bayesian inverse problems with Gaussian priors. *The Annals of Statistics*, 39(5):2626–2657, 2011.
- [36] V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen. Sparsity-promoting bayesian inversion. *Inverse Problems*, 28(2):025005, 2012.
- [37] M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging*, 3(1):87–122, 2009.
- [38] M. Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2005.
- [39] P. J. Lenk. The logistic normal distribution for bayesian, nonparametric, predictive densities. *Journal of the American Statistical Association*, 83(402):509–516, 1988.
- [40] P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.
- [41] Y. Meyer. Wavelets and operators. volume 37 of *Cambridge Studies in Advanced Mathematics*, 1992.
- [42] O. Papaspiliopoulos, Y. Pokern, G. O. Roberts, and A. M. Stuart. Non-parametric estimation of diffusions: a differential equations approach. *Biometrika*, 99(3):511–531, 2012.
- [43] Y. Pokern, A. M. Stuart, and J. H. van Zanten. Posterior consistency via

- precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.*, 123(2):603–628, 2013.
- [44] K. Ray et al. Bayesian inverse problems with non-conjugate priors. *Electronic Journal of Statistics*, 7:2516–2549, 2013.
- [45] C. Roberto. Isoperimetry for product of probability measures: recent results. *Markov Process. Related Fields*, 16(4):617–634, 2010.
- [46] J. Rousseau and B. Szabo. Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator. *Ann. Statist.*, 45(2):833–865, 2017.
- [47] H. J. Schmeisser and H. Triebel. *Topics in Fourier analysis and function spaces*, volume 42 of *Mathematics and its Applications in Physics and Technology*. Akademische Verlagsgesellschaft Geest & Portig K. G., Leipzig, 1987.
- [48] X. Shen and L. Wasserman. Rates of convergence of posterior distributions. *The Annals of Statistics*, 29(3):687–714, 2001.
- [49] L. Shepp. Distinguishing a sequence of random variables from a translate of itself. *The Annals of Mathematical Statistics*, 36(4):1107–1112, 1965.
- [50] T. S. Shively, R. Kohn, and S. Wood. Variable selection and function estimation in additive nonparametric regression using a data-based prior. *Journal of the American Statistical Association*, 94(447):777–794, 1999.
- [51] W. Stolz. Some small ball probabilities for Gaussian processes under nonuniform norms. *Journal of Theoretical Probability*, 9(3):613–630, 1996.
- [52] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.
- [53] B. T. Szabó, A. W. van der Vaart, and J. H. van Zanten. Empirical Bayes scaling of Gaussian priors in the white noise model. *Electronic Journal of Statistics*, 7:991–1018, 2013.
- [54] M. Talagrand. The supremum of some canonical processes. *American Journal of Mathematics*, 116(2):283–325, 1994.
- [55] H. Triebel. *Interpolation theory, function spaces, differential operators*, volume 18 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam-New York, 1978.
- [56] H. Triebel. *Bases in function spaces, sampling, discrepancy, numerical integration*, volume 11. European Mathematical Society, 2010.
- [57] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- [58] A. W. Van Der Vaart and J. H. van Zanten. Bayesian inference with rescaled gaussian process priors. *Electronic Journal of Statistics*, 1:433–448, 2007.
- [59] A. W. van der Vaart and J. H. van Zanten. Rates of contraction of posterior distributions based on gaussian process priors. *The Annals of Statistics*, pages 1435–1463, 2008.
- [60] A. W. van der Vaart and J. H. van Zanten. Reproducing kernel Hilbert spaces of Gaussian priors. pages 200–222, 2008.
- [61] A. W. van der Vaart and J. H. van Zanten. Adaptive bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *The Annals*

of Statistics, 37(5B):2655–2675, 2009.

- [62] L. H. Zhao. Bayesian aspects of some nonparametric problems. *Annals of statistics*, pages 532–552, 2000.

SUPPLEMENTARY MATERIAL

Note that in order to ease readability, in this supplementary material we use a different type of numbering for sections, results and displayed equations, compared to the main body of the article. In particular, we use letters for sections, and the letter of the section together with number, for results and displayed equations.

A. Proofs of Section 2

Proof of Proposition 2.10. By Proposition 2.9 we have $\overline{\mathcal{Z}}^{\|\cdot\|_X} \subset \overline{\mathcal{Q}}^{\|\cdot\|_X} \subset X$. For any arbitrary $x \in X$ and given $\epsilon > 0$, there exists N such that $x^N = \sum_{\ell=1}^N x_\ell \psi_\ell$ satisfies $\|x^N - x\|_X < \epsilon$, where (ψ_ℓ) is the Schauder basis in X and (x_ℓ) the corresponding coefficients of x . Since clearly $x^N \in \mathcal{Q} \cap \mathcal{Z}$, we conclude that $X = \overline{\mathcal{Z}}^{\|\cdot\|_X} = \overline{\mathcal{Q}}^{\|\cdot\|_X}$.

Since μ is a measure on X , we have $\text{supp}(\mu) \subset X$. On the other hand, the topological support of any Radon measure in X is non-empty and by definition closed in X . By Proposition 2.4 we get that $0 \in \text{supp}(\mu)$, thus $\mathcal{Q} \subset \text{supp}(\mu)$. Taking closures in X , we get $X = \overline{\mathcal{Q}}^{\|\cdot\|_X} \subset \text{supp}(\mu)$ and thus the claimed result. \square

Proof of Proposition 2.11. By Proposition 2.7, letting $V = \frac{|\cdot|^p}{p}$, we have

$$\begin{aligned} \mu(\epsilon B_X + h) &= \int_{\epsilon B_X} \lim_{N \rightarrow \infty} e^{\left(\sum_{\ell=1}^N \left(V\left(\frac{u_\ell}{\gamma_\ell}\right) - V\left(\frac{u_\ell - h_\ell}{\gamma_\ell}\right)\right)\right)} \mu(du) \\ &= e^{-\sum_{\ell=1}^\infty V\left(\frac{h_\ell}{\gamma_\ell}\right)} \int_{\epsilon B_X} \lim_{N \rightarrow \infty} e^{\sum_{\ell=1}^N \left(V\left(\frac{u_\ell}{\gamma_\ell}\right) + V\left(\frac{h_\ell}{\gamma_\ell}\right) - V\left(\frac{u_\ell - h_\ell}{\gamma_\ell}\right)\right)} \mu(du) \\ &= e^{-\sum_{\ell=1}^\infty V\left(\frac{h_\ell}{\gamma_\ell}\right)} \int_{\epsilon B_X} \lim_{N \rightarrow \infty} \frac{1}{2} \left(e^{\sum_{\ell=1}^N \left(V\left(\frac{u_\ell}{\gamma_\ell}\right) + V\left(\frac{h_\ell}{\gamma_\ell}\right) - V\left(\frac{u_\ell - h_\ell}{\gamma_\ell}\right)\right)} \right. \\ &\quad \left. + e^{\sum_{\ell=1}^N \left(V\left(\frac{u_\ell}{\gamma_\ell}\right) + V\left(\frac{h_\ell}{\gamma_\ell}\right) - V\left(\frac{u_\ell + h_\ell}{\gamma_\ell}\right)\right)} \right) \mu(du), \end{aligned}$$

where in the last equality we used symmetry. In the following, we show that the integrand in the last line above is bounded below by 1. Notice that our proof applies for any $V : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ convex and symmetric with $V(0) = 0$, differentiable in $\mathbb{R} \setminus \{0\}$ with concave derivative on the positive axis. The functions $\frac{|\cdot|^p}{p}$, $1 \leq p \leq 2$, clearly satisfy this assumption.

Since $e^a + e^{-a} \geq 2$ for any $a \in \mathbb{R}$, we observe that

$$\begin{aligned} &e^{V(x)+V(y)-V(x-y)} + e^{V(x)+V(y)-V(x+y)} \\ &= e^{V(x)+V(y)-\frac{1}{2}V(x-y)-\frac{1}{2}V(x+y)} \left(e^{-\frac{1}{2}V(x-y)+\frac{1}{2}V(x+y)} + e^{\frac{1}{2}V(x-y)-\frac{1}{2}V(x+y)} \right) \\ &\geq 2e^{V(x)+V(y)-\frac{1}{2}V(x-y)-\frac{1}{2}V(x+y)}. \end{aligned}$$

In consequence, we need to show that

$$G(x, y) := V(x) + V(y) - \frac{1}{2}V(x - y) - \frac{1}{2}V(x + y) \geq 0.$$

Notice that G has a number of symmetries. Namely, it satisfies

$$G(x, y) = G(-x, y) = G(x, -y) = G(y, x). \quad (\text{A.1})$$

Therefore, without loss of generality, we can assume that $x, y \geq 0$. We note that if $x = 0$ or $y = 0$ then clearly $G(x, y) = V(0) = 0$. Consequently, due to (A.1) it will be sufficient to show that $\frac{\partial G}{\partial x}(x, y) \geq 0$ for any $x > 0$ and $y > 0$.

Let us briefly consider the derivative $R(x) = V'(x)$ for $x > 0$ and define $R(0) = \lim_{x \rightarrow 0^+} V'(x)$. By assumption on V , $V'(x)$ is concave hence continuous for all $x > 0$, implying that the limit exists although it may be $-\infty$. Combining with the convexity of V and since V has a minimum at the origin, we get that the limit is non-negative, $R(0) \geq 0$. The function R defined on $[0, \infty)$ is concave with $R(0) \geq 0$, hence it is subadditive.

We first observe that

$$G(x, x) = 2V(x) - \frac{1}{2}V(2x)$$

and due to the subadditivity of R , we must have $G(x, x) \geq 0$. For $x \neq y$, we have

$$\frac{\partial G}{\partial x}(x, y) = V'(x) - \frac{1}{2}V'(x - y) - \frac{1}{2}V'(x + y).$$

For $x > y$, by concavity of V' on the positive axis, we have

$$V'(x) = V'\left(\frac{1}{2}(x - y) + \frac{1}{2}(x + y)\right) \geq \frac{1}{2}V'(x - y) + \frac{1}{2}V'(x + y),$$

implying that $\frac{\partial G}{\partial x}(x, y) \geq 0$ for $x \geq y$. If $x < y$, since by symmetry of V it holds $V'(x - y) = -V'(y - x)$, we can write

$$\frac{\partial G}{\partial x}(x, y) = V'(x) + \frac{1}{2}V'(y - x) - \frac{1}{2}V'(x + y)$$

where the arguments of V' in the right-hand side are positive and we can use the concavity of V' on the positive axis. As above, using the auxiliary function R and since concave functions which are non-negative at zero are subadditive, we have

$$V'(x + y) = V'(2x + y - x) \leq 2V'(x) + V'(y - x).$$

Thus $\frac{dG}{dx}(x; y) \geq 0$ for any $y > x > 0$ as well, and the result follows. \square

Proof of Theorem 2.13. Let $h \in \mathcal{Z}$ such that $\|h - w\|_X \leq \epsilon$. Then by the triangle inequality, for any $x \in X$ we have $\|x - w\|_X \leq \epsilon + \|x - h\|_X$, hence if $\|x - h\|_X \leq \epsilon$ then $\|x - w\|_X \leq 2\epsilon$. We thus have,

$$\mu(w + 2\epsilon B_X) \geq \mu(h + \epsilon B_X) \geq e^{-\frac{1}{p}\|h\|_{\mathcal{Z}}^p} \mu(\epsilon B_X),$$

where for the last inequality we used Proposition 2.11. To finish the proof we take the negative logarithm and optimize over $h \in \mathcal{Z}$. \square

Proof of Proposition 2.15. Without loss of generality we work in \mathbb{R}^∞ . Recall $\gamma = (\gamma_\ell)$ and $\xi = (\xi_\ell)$ from the definition of the p -exponential measure μ , Definition 2.1. The inequality follows from [54, Theorem 2.4], see also [38, Theorem 4.19]. These theorems state that for the infinite (unscaled) independent product of standard p -exponential one-dimensional measures, μ_∞ in \mathbb{R}^∞ , there exists a universal constant $K > 0$ depending only on p , such that for all $\tilde{r} > 0$

$$\mu_\infty(A + \sqrt{\tilde{r}}B_2 + \tilde{r}^{\frac{1}{p}}B_p) \geq 1 - \frac{1}{\mu_\infty(A)} \exp\left(-\frac{\tilde{r}}{K}\right),$$

where B_p, B_2 are the closed unit balls in ℓ_p, ℓ_2 respectively. Letting $r = \tilde{r}^{\frac{1}{p}}$, we get

$$\mu_\infty(A + r^{\frac{p}{2}}B_2 + rB_p) \geq 1 - \frac{1}{\mu_\infty(A)} \exp\left(-\frac{r^p}{K}\right).$$

Defining $\Gamma : \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$, such that $x \in \mathbb{R}^\infty \mapsto (\gamma_\ell x_\ell)$, we get that for any μ -measurable set $A \subset \mathbb{R}^\infty$

$$\begin{aligned} \mu(A + r^{\frac{p}{2}}B_Q + rB_Z) &= \mathbb{P}(\Gamma\xi \in A + r^{\frac{p}{2}}B_Q + rB_Z) = \mathbb{P}(\xi \in \Gamma^{-1}(A + r^{\frac{p}{2}}B_Q + rB_Z)) \\ &= \mu_\infty(\Gamma^{-1}A + r^{\frac{p}{2}}B_2 + rB_p) \geq 1 - \frac{1}{\mu_\infty(\Gamma^{-1}A)} \exp\left(-\frac{r^p}{K}\right) = 1 - \frac{1}{\mu(A)} \exp\left(-\frac{r^p}{K}\right). \end{aligned}$$

□

B. Proof of general contraction theorem in Section 3

Proof of Theorem 3.1. Assume $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$. It follows by Theorem 2.13 that

$$\mathbb{P}(\|W - w_0\|_X < 2\epsilon_n) = \exp(\log \mu(w_0 + 2\epsilon_n B_X)) \geq \exp(-\varphi_{w_0}(\epsilon_n)) \geq e^{-n\epsilon_n^2},$$

and, consequently, the claim (9) follows.

We now consider the existence of sets X_n such that (7) and (8) hold. We set

$$X_n = \epsilon_n B_X + M_n^{\frac{p}{2}} B_Q + M_n B_Z, \quad (\text{B.2})$$

where $M_n > 0$ will be chosen below. By Proposition 2.15, we have

$$\mathbb{P}(W \notin X_n) \leq \frac{1}{\mu(\epsilon_n B_X)} \exp\left(-\frac{M_n^p}{K}\right) = \exp\left(\varphi_0(\epsilon_n) - \frac{M_n^p}{K}\right). \quad (\text{B.3})$$

Next, for any $C > 1$, we denote $M_n = (K(C+1)n\epsilon_n^2)^{\frac{1}{p}}$ which is bounded away from zero for all n by assumption. Since

$$\varphi_0(\epsilon_n) \leq \varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2 \quad (\text{B.4})$$

we obtain the claim (8) by combining (B.3) with (B.4).

For the final claim (7), we cannot use directly Proposition 2.11 to bound the complexity of X_n , since Proposition 2.11 refers to shifts in \mathcal{Z} while X_n involves

a ball in \mathcal{Q} . We can however, find a large enough ball $\overline{M}_n \mathcal{Z}$ which is such that a $2\epsilon_n$ -cushion in X around it contains X_n . We can then use Proposition 2.11 to bound the complexity of $2\epsilon_n B_X + \overline{M}_n B_{\mathcal{Z}}$, which in turn implies a bound on the complexity of X_n .

Define

$$\overline{M}_n = 2 \left(M_n \vee \left(1 + \frac{1}{n}\right) f(M_n^{\frac{p}{2}})^{\frac{1}{p}} g(\epsilon_n)^{\frac{1}{p} - \frac{1}{2}} \right).$$

Then using (6) we can show that

$$X_n \subset 2\epsilon_n B_X + \overline{M}_n B_{\mathcal{Z}}. \quad (\text{B.5})$$

Indeed, for every $x \in M_n^{\frac{p}{2}} B_{\mathcal{Q}}$, we have by (6) that

$$\inf_{z \in \mathcal{Z} : \|z - x\|_X \leq \epsilon} \|z\|_{\mathcal{Z}}^p \leq f(M_n^{\frac{p}{2}}) g(\epsilon)^{1 - \frac{p}{2}}$$

and, in consequence, there exists $y \in \left(1 + \frac{1}{n}\right) f(M_n^{\frac{p}{2}})^{\frac{1}{p}} g(\epsilon_n)^{\frac{1}{p} - \frac{1}{2}} B_{\mathcal{Z}}$, with $\|x - y\|_X \leq \epsilon_n$. The term $1 + 1/n$ does not play any significant role, and can be replaced by any constant over 1. Thus any $x \in M_n^{\frac{p}{2}} B_{\mathcal{Q}} + M_n B_{\mathcal{Z}}$ is within $\epsilon_n \|\cdot\|_X$ -distance from some point in $\overline{M}_n B_{\mathcal{Z}}$ and (B.5) follows.

Let $h_1, \dots, h_N \in \overline{M}_n B_{\mathcal{Z}}$ be $2\epsilon_n$ -apart in $\|\cdot\|_X$. Clearly, the balls $h_j + \epsilon_n B_X$ are disjoint and hence by Proposition 2.11 we obtain

$$1 \geq \sum_{j=1}^N \mathbb{P}(W \in h_j + \epsilon_n B_X) \geq \sum_{j=1}^N e^{-\frac{\|h_j\|_{\mathcal{Z}}^p}{p}} \mathbb{P}(W \in \epsilon_n B_X) \geq N e^{-\frac{\overline{M}_n^p}{p} - \varphi_0(\epsilon_n)}. \quad (\text{B.6})$$

If the set of points h_1, \dots, h_N is maximal in $\overline{M}_n B_{\mathcal{Z}}$ (that is, it achieves the maximum number of points $2\epsilon_n$ -apart in $\|\cdot\|_X$ that can fit in $\overline{M}_n B_{\mathcal{Z}}$), then the balls $h_j + 2\epsilon_n B_X$ cover $\overline{M}_n B_{\mathcal{Z}}$ and combining with (B.5) we get that

$$X_n \subset \bigcup_{j=1}^N (h_j + 4\epsilon_n B_X). \quad (\text{B.7})$$

Combining (B.6) together with (B.7) we obtain

$$N(4\epsilon_n, X_n, \|\cdot\|_X) \leq N \leq \exp \left(\frac{\overline{M}_n^p}{p} + \varphi_0(\epsilon_n) \right).$$

Using the definitions of M_n, \overline{M}_n we get

$$\begin{aligned} & \log N(4\epsilon_n, X_n, \|\cdot\|_X) \\ & \leq \frac{2^p}{p} \left(K(C+1)n\epsilon_n^2 \vee \left(1 + \frac{1}{n}\right)^p f(\sqrt{K(C+1)n^{\frac{1}{2}}\epsilon_n}) g(\epsilon_n)^{1 - \frac{p}{2}} \right) + n\epsilon_n^2. \end{aligned}$$

Finally, using that $n\epsilon_n^2 \gtrsim 1$ and f is non-decreasing with $f(a) \rightarrow \infty$ at most polynomially as $a \rightarrow \infty$, we get (7). This completes the proof. \square

C. Proofs of Section 5

Proof of Lemma 5.2. The proof is very similar to the proof of [17, Theorem 5], taking into account Proposition 2.5. We include it for the reader's convenience.

We first show that for $s < \alpha$, it holds $\mu(B_q^s) = 1$. Indeed, we have

$$\mathbb{E}_\mu \|u\|_{B_q^s}^q = \mathbb{E} \sum_{\ell=1}^{\infty} \ell^{q(\frac{s}{d} + \frac{1}{2})-1} |\gamma_\ell|^q |\xi_\ell|^q = \mathbb{E} |\xi_1|^q \sum_{\ell=1}^{\infty} \ell^{\frac{q(s-\alpha)}{d}-1}.$$

If $s < \alpha$, then the expectation is finite, hence $\mu(B_q^s) = 1$.

We next show that if $\mu(B_q^s) = 1$ then $s < \alpha$. If $\mu(B_q^s) = 1$ then $\|u\|_{B_q^s} < \infty$ almost surely with respect to μ , hence

$$\sum_{\ell=1}^{\infty} \ell^{\frac{q(s-\alpha)}{d}-1} |\xi_\ell|^q < \infty, \quad \text{almost surely.} \quad (\text{C.8})$$

By contradiction to the Law of Large Numbers, we then get that

$$s < \alpha + \frac{d}{q}. \quad (\text{C.9})$$

Define $\zeta_\ell = \ell^{\frac{q(s-\alpha)}{d}-1} |\xi_\ell|^q$, which are independent non-negative random variables. By [32, Proposition 4.14], (C.8) also implies that

$$\sum_{\ell=1}^{\infty} \mathbb{E}[\zeta_\ell \wedge 1] < \infty. \quad (\text{C.10})$$

We have

$$\mathbb{E}\zeta_\ell = \mathbb{E}[\zeta_\ell \mathbb{1}_{\zeta_\ell \leq 1}] + \mathbb{E}[\zeta_\ell \mathbb{1}_{\zeta_\ell > 1}] \leq \mathbb{E}[\zeta_\ell \wedge 1] + I_\ell,$$

where

$$I_\ell = \mathbb{E} \left[\ell^{\frac{q(s-\alpha)}{d}-1} |\xi_\ell|^q \mathbb{1}_{\{|\xi_\ell| > \ell^{\frac{\alpha-s}{d} + \frac{1}{q}}\}} \right] = c_p \ell^{\frac{q(s-\alpha)}{d}-1} \int_{\ell^{\frac{\alpha-s}{d} + \frac{1}{q}}}^{\infty} x^q e^{-\frac{x^p}{p}} dx,$$

where c_p depends only on p . Since $p, q \geq 1$, it holds that $x^q e^{-\frac{x^p}{p}} \leq C_1 e^{-C_2 x}$, for constants $C_1, C_2 > 0$ sufficiently large and small, respectively. This results in the bound

$$I_\ell \leq c_p \frac{C_1}{C_2} \ell^{\frac{q(s-\alpha)}{d}-1} \exp(-C_2 \ell^{\frac{\alpha-s}{d} + \frac{1}{q}}) := \iota_\ell,$$

where by (C.9) ι_ℓ are summable. Combining, we get that

$$\sum_{\ell=1}^{\infty} \mathbb{E}[\zeta_\ell] \leq \sum_{\ell=1}^{\infty} \mathbb{E}[\zeta_\ell \wedge 1] + \sum_{\ell=1}^{\infty} \iota_\ell < \infty.$$

We thus have

$$\sum_{\ell=1}^{\infty} \mathbb{E}[\zeta_{\ell}] = \mathbb{E}[|\xi_1|^q] \sum_{\ell=1}^{\infty} \ell^{\frac{q(s-\alpha)}{d}-1} < \infty,$$

and therefore $s < \alpha$.

Finally, Proposition 2.5 implies that $\mu(B_q^s) = 0$, for $s \geq \alpha$. \square

Proof of Proposition 5.4. We examine each case separately.

- i) For $q \geq 2$, by Lemmas 5.12 and 5.13, we have that as $\epsilon \rightarrow 0$ the concentration function satisfies

$$\varphi_{w_0}(\epsilon) \lesssim \begin{cases} \epsilon^{\frac{\beta p - \alpha p - d}{\beta}} + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta < \alpha + \frac{d}{p}, \\ (-\log \epsilon)^{\frac{q-p}{q}} + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta = \alpha + \frac{d}{p}, \\ 1 + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta > \alpha + \frac{d}{p}. \end{cases} \quad (\text{C.11})$$

Determining which term dominates in each case, we find

$$\varphi_{w_0}(\epsilon) \lesssim \begin{cases} \epsilon^{\frac{\beta p - \alpha p - d}{\beta}}, & \text{for } \alpha \geq \beta, \\ \epsilon^{-\frac{d}{\alpha}}, & \text{for } \alpha < \beta. \end{cases} \quad (\text{C.12})$$

Computing the minimal solution ϵ_n such that $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$, we arrive at $\epsilon_n \asymp r_n^{\alpha, \beta, p, q}$.

- ii) For $q < 2$ and $p \leq q$, by Lemmas 5.12 and 5.13, we have that as $\epsilon \rightarrow 0$ the concentration function satisfies

$$\varphi_{w_0}(\epsilon) \lesssim \begin{cases} \epsilon^{\frac{2\beta pq - 2\alpha pq - 2qd}{2\beta q + qd - 2d}} + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta < \alpha + \frac{d}{p}, \\ (-\log \epsilon)^{\frac{q-p}{q}} + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta = \alpha + \frac{d}{p}, \\ 1 + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta > \alpha + \frac{d}{p}. \end{cases} \quad (\text{C.13})$$

In this case, determining which of the two terms dominates for $\beta < \alpha + \frac{d}{p}$ is a bit more complicated and leads to a quadratic equation for the value of α balancing the two terms. This quadratic equation has two solutions, a negative one which is rejected since $\alpha > 0$ and $\alpha = \frac{\beta p - d + a}{2p}$, where a as in the statement of the lemma. Notice that the expression under the square root in a is positive by the assumption on β . We thus have the following bound:

$$\varphi_{w_0}(\epsilon) \lesssim \begin{cases} \epsilon^{\frac{2\beta pq - 2\alpha pq - 2qd}{2\beta q + qd - 2d}}, & \text{for } \alpha \geq \frac{\beta p - d + a}{2p}, \\ \epsilon^{-\frac{d}{\alpha}}, & \text{for } \alpha < \frac{\beta p - d + a}{2p}. \end{cases} \quad (\text{C.14})$$

Computing the minimal solution ϵ_n such that $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$, we again arrive at $\epsilon_n \asymp r_n^{\alpha, \beta, p, q}$.

- iii) For $q < 2$, thus $p > q$, the proof is similar to (ii) above, using the expressions corresponding to this case from Lemma 5.13.

Finally, notice that the constants in all the used upper bounds on the concentration function from Lemmas 5.12 and 5.13, depend on w_0 only through its B_q^β norm, hence so do the constants in the rates derived above. \square

Proof of Proposition 5.8. In all of the studied cases except when $\alpha = \beta - \frac{d}{p}$ for $q > p$, Lemmas 5.12 and 5.13 result in a bound of the form

$$\bar{\varphi}_{w_0}(\epsilon) \lesssim \lambda^{-p} \epsilon^{-s} + (\epsilon/\lambda)^{-\frac{d}{\alpha}},$$

for some $s = s(\alpha, \beta, p, q) \geq 0$. We optimize the choice of λ by balancing the two terms, obtaining the choice

$$\lambda \asymp \epsilon^{\frac{d-\alpha s}{d+\alpha p}}. \quad (\text{C.15})$$

We then equate the resulting bound on $\bar{\varphi}_{w_0}(\epsilon)$ with $n\epsilon^2$, thus getting

$$\epsilon_n \asymp n^{-(2+\frac{dp+ds}{d+\alpha p})^{-1}}. \quad (\text{C.16})$$

The last rate coincides with the rate $m_n = n^{-\frac{\beta}{d+2\beta}}$, if and only if

$$s = s_0(\alpha, \beta, p) := \frac{d + \alpha p - \beta p}{\beta}.$$

The corresponding optimal λ is

$$\lambda_n \asymp n^{\frac{\alpha-\beta}{d+2\beta}}. \quad (\text{C.17})$$

For $s > s_0$, the rate ϵ_n is polynomially slower than m_n , and this holds for the optimal and hence any choice of λ_n . The case $s < s_0$ does not arise, as seen below (it cannot arise as it would lead to a faster rate than the minimax rate m_n). The proof thus proceeds by comparing the values of s obtained in Lemma 5.13 to s_0 .

Recall that $q < 2$. If $q \leq p$ and $\alpha \geq \beta - \frac{d}{q}$, we have

$$s = \frac{2\alpha pq + 2pd - 2\beta pq}{2\beta q + dq - 2d}. \quad (\text{C.18})$$

It holds

$$s - s_0 = \frac{dq(\beta - \alpha - d/q)(p-2) + (d^2 + 2\alpha d)(p-q)}{\beta(2\beta q + dq - 2d)} \geq 0,$$

since both numerator and denominator are positive by our assumptions. The difference $s - s_0$ vanishes if and only if $\alpha = \beta - \frac{d}{q}$ and $q = p$, in which case $\epsilon_n \asymp m_n$ for $\lambda_n \asymp n^{-\frac{d}{p(d+2\beta)}}$. In all other cases, for any λ_n , ϵ_n is polynomially slower than m_n .

If $\alpha < \beta - \frac{d}{p}$ for any relationship between q and p , we have $s = 0 > s_0$, hence for any λ_n , ϵ_n is polynomially slower than m_n .

If $q > p$ and $\alpha > \beta - \frac{d}{p}$, then

$$s = \frac{2\alpha pq + 2qd - 2\beta pq}{2\beta q + qd - 2d}.$$

It holds

$$s - s_0 = \frac{(\beta - \alpha - \frac{d}{p})(q - 2)dp}{\beta(2\beta q + qd - 2d)} > 0,$$

hence for any λ_n , ϵ_n is polynomially slower than m_n .

We next turn to the case $q > p$ for $\alpha = \beta - \frac{d}{p}$ (recall $q < 2$). By Lemmas 5.12 and 5.13 we have the bound

$$\bar{\varphi}_{w_0}(\epsilon) \lesssim \lambda^{-p}(\log 1/\epsilon)^{\frac{q-p}{q}} + (\epsilon/\lambda)^{-\frac{d}{\alpha}}.$$

We again optimize the choice of λ by balancing the two terms, to find

$$\lambda \asymp \epsilon^{\frac{d}{\beta p}} (\log 1/\epsilon)^{\frac{q-p}{q} \frac{\beta p - d}{\beta p^2}}. \quad (\text{C.19})$$

The resulting bound on $\bar{\varphi}_{w_0}(\epsilon)$ is

$$\bar{\varphi}_{w_0}(\epsilon) \lesssim (\log 1/\epsilon)^{\frac{d(q-p)}{\beta q p}} \epsilon^{-\frac{d}{\beta}}.$$

We equate this bound with $n\epsilon^2$, obtaining the equation

$$\epsilon_n^{\frac{d+2\beta}{\beta}} \log^{-\frac{(q-p)d}{\beta q p}}(1/\epsilon_n) = n^{-1}.$$

We then solve for ϵ using [40, Lemma 3], included below as Lemma G.2 for the reader's convenience. This gives the claimed value of \bar{r}_n and we can in turn compute the value of λ_n , by plugging $\epsilon_n \asymp \bar{r}_n$ in (C.19). It holds that $\omega > 0$ (see the expression for λ_n in the statement), since $\alpha = \beta - \frac{d}{p} > 0$.

Finally, we return to the case $q < p$ and determine the best achievable rate. If $\alpha \leq \beta - \frac{d}{q}$, then $s = 0$ and by (C.16), the rate is

$$\epsilon_n \asymp n^{-(2 + \frac{dp}{d + \alpha p})^{-1}},$$

which is (uniquely) optimized when we choose α as large as possible, $\alpha = \beta - \frac{d}{q}$. Plugging this choice of α into the last expression and (C.15), we obtain the claimed values for \bar{r}_n and λ_n , respectively. It remains to verify that for $\alpha > \beta - \frac{d}{q}$ the resulting rate ϵ_n in (C.16) is not better than the aforementioned \bar{r}_n . Indeed, comparing ϵ_n and \bar{r}_n , we conclude that this is the case if and only if the following inequality holds

$$s \geq p^2 \frac{\alpha q + d - \beta q}{dq - dp + p\beta q},$$

with equality if and only if the two rates coincide. Using the expression for s from (C.18), we get that the last inequality is equivalent to

$$\frac{2}{2\beta q + dq - 2d} \geq \frac{p}{\beta p q + dq - dp},$$

where notice that the two denominators are positive since $\beta \geq \frac{d}{q} - \frac{d}{2}$ and $p \leq 2$. It is then straightforward to see that the last inequality is equivalent to $p \leq 2$

and so indeed the rate for $\alpha > \beta - \frac{d}{q}$ is strictly worse than \bar{r}_n when $p < 2$, while for $p = 2$ the rates are identical for any $\alpha \geq \beta - \frac{d}{q}$.

As always, notice that the constants in all the used upper bounds on the concentration function from Lemmas 5.12 and 5.13, depend on w_0 only through its B_q^β norm, hence so do the constants in the rates derived above. \square

Proof of Lemma 5.13. Let $w_0 := (w_{0,\ell})_{\ell \in \mathbb{N}} \in B_q^\beta$, so that under the assumption on β it holds that $w_0 \in \ell_2$, see Lemma G.1 below. Since $w_0 \in B_q^\beta$ we find that

$$|w_{0,\ell}| \leq \|w_0\|_{B_q^\beta} \ell^{-\frac{\beta}{d} - \frac{1}{2} + \frac{1}{q}}. \quad (\text{C.20})$$

Consider now approximations $h_{1:L} = (w_1, \dots, w_L, 0, \dots) \in \mathbb{R}^\infty$ of w_0 , where $L \in \mathbb{N}$. Obviously, we have $h_{1:L} \in \mathcal{Z}_\alpha$ for any L . We first study how large L needs to be, in order to have

$$\|h_{1:L} - w_0\|_{\ell_2} \leq \epsilon. \quad (\text{C.21})$$

In case $q > 2$, we have

$$\begin{aligned} \|h_{1:L} - w_0\|_{\ell_2}^2 &= \sum_{\ell > L} \ell^{-\frac{2\beta}{d} - 1 + \frac{2}{q}} \ell^{\frac{2\beta}{d} + 1 - \frac{2}{q}} w_{0,\ell}^2 \\ &\leq \left(\sum_{\ell > L} \ell^{\frac{\beta}{d} + \frac{q}{2} - 1} |w_{0,\ell}|^q \right)^{\frac{2}{q}} \left(\sum_{\ell > L} \ell^{-\frac{2\beta}{d(q-2)} - 1} \right)^{\frac{q-2}{q}} \leq \|w_0\|_{B_q^\beta}^2 L^{-\frac{2\beta}{d}}, \end{aligned}$$

where we have used the Hölder inequality $(\frac{q}{2}, \frac{q}{q-2})$ and comparison of the sum to an integral. If $q = 2$, we obtain

$$\|h_{1:L} - w_0\|_{\ell_2}^2 = \sum_{\ell > L} \ell^{\frac{2\beta}{d}} w_{0,\ell}^2 \ell^{-\frac{2\beta}{d}} \leq L^{-\frac{2\beta}{d}} \|w_0\|_{B_q^\beta}^2.$$

If $q < 2$, applying (C.20) yields

$$\begin{aligned} \|h_{1:L} - w_0\|_{\ell_2}^2 &= \sum_{\ell > L} |w_{0,\ell}|^q |w_{0,\ell}|^{2-q} \leq \|w_0\|_{B_q^\beta}^{2-q} \sum_{\ell > L} |w_{0,\ell}|^q \ell^{\frac{\beta}{d} + \frac{q}{2} - 1} \ell^{-\frac{2\beta}{d} - 1 + \frac{2}{q}} \\ &\leq L^{\frac{2}{q} - 1 - \frac{2\beta}{d}} \|w_0\|_{B_q^\beta}^2, \end{aligned}$$

where we used the assumption on β and where all the constants depend on w_0 only through its B_q^β -norm.

For (C.21) to hold with minimal $L \in \mathbb{N}$, we choose L as

$$L = c \begin{cases} \epsilon^{-\frac{d}{\beta}}, & \text{if } q \geq 2, \\ \epsilon^{\frac{1}{\frac{1}{q} - \frac{1}{2} - \frac{\beta}{d}}}, & \text{if } q < 2, \end{cases} \quad (\text{C.22})$$

where the constant c depends on w_0 only through its B_q^β -norm.

Testing $h_{1:L} \in \mathcal{Z}_\alpha$ in the infimum we aim to bound, yields an upper bound

$$I(\epsilon) := \inf_{h \in \mathcal{Z}_\alpha : \|h - w_0\|_{\ell_2} \leq \epsilon} \|h\|_{\mathcal{Z}_\alpha}^p \leq \|h_{1:L}\|_{\mathcal{Z}_\alpha}^p = \sum_{\ell \leq L} \ell^{\frac{p}{2} + \frac{\alpha p}{d}} |w_{0,\ell}|^p.$$

i) For $q \leq p$, we use (C.20) to get

$$\begin{aligned} \|h_{1:L}\|_{\mathcal{Z}_\alpha}^p &= \sum_{\ell \leq L} \ell^{\frac{p}{2} + \frac{\alpha p}{d}} |w_{0,\ell}|^p = \sum_{\ell \leq L} \ell^{\frac{\beta q}{d} + \frac{q}{2} - 1} |w_{0,\ell}|^q |w_{0,\ell}|^{p-q} \ell^{\frac{p}{2} + \frac{\alpha p}{d} - \frac{\beta q}{d} - \frac{q}{2} + 1} \\ &\leq \|w_0\|_{B_q^\beta}^{p-q} \sum_{\ell \leq L} \ell^{\frac{\beta q}{d} + \frac{q}{2} - 1} |w_{0,\ell}|^q \ell^{\frac{(\alpha-\beta)p}{d} + \frac{p}{q}} \\ &\leq \begin{cases} \|w_0\|_{B_q^\beta}^p, & \text{if } \beta \geq \alpha + \frac{d}{q}, \\ \|w_0\|_{B_q^\beta}^p L^{\frac{(\alpha-\beta)p}{d} + \frac{p}{q}}, & \text{if } \beta < \alpha + \frac{d}{q}. \end{cases} \end{aligned} \quad (\text{C.23})$$

ii) For $q > p$, we use Hölder inequality with $(\frac{q}{p}, \frac{q}{q-p})$ to bound

$$\begin{aligned} \|h_{1:L}\|_{\mathcal{Z}_\alpha}^p &= \sum_{\ell \leq L} \ell^{\frac{p}{2} + \frac{\alpha p}{d}} |w_{0,\ell}|^p = \sum_{\ell \leq L} \ell^{\frac{\beta p}{d} + \frac{p}{2} - \frac{p}{q}} |w_{0,\ell}|^p \ell^{\frac{\alpha p}{d} - \frac{\beta p}{d} + \frac{p}{q}} \\ &\leq \left(\sum_{\ell \leq L} \ell^{\frac{\beta q}{d} + \frac{q}{2} - 1} |w_{0,\ell}|^q \right)^{\frac{p}{q}} \left(\sum_{\ell \leq L} \ell^{(\frac{\alpha p}{d} - \frac{\beta p}{d} + \frac{p}{q}) \frac{q}{q-p}} \right)^{\frac{q-p}{q}} \\ &\leq \begin{cases} \|w_0\|_{B_q^\beta}^p, & \text{if } \beta \geq \alpha + \frac{d}{p}, \\ \|w_0\|_{B_q^\beta}^p (\log L)^{\frac{q-p}{q}}, & \text{if } \beta = \alpha + \frac{d}{p}, \\ \|w_0\|_{B_q^\beta}^p L^{\frac{\alpha p - \beta p + d}{d}}, & \text{if } \beta < \alpha + \frac{d}{p}. \end{cases} \end{aligned} \quad (\text{C.24})$$

Combining the bounds (C.23) and (C.24) with the choice of L from (C.22) completes the proof. \square

Proof of Lemma 5.14. We first show that f and g defined by (18) satisfy (6). Let $h = (h_\ell) \in aB_{\mathcal{Q}_\alpha}$ and define $x_{1:L} = (h_1, \dots, h_L, 0, \dots)$ for the smallest $L = L(\epsilon; a) \in \mathbb{N}$ such that $\gamma_L \leq \frac{\epsilon}{a}$. Such L satisfies

$$L = \lceil a^{\frac{2d}{d+2\alpha}} \epsilon^{-\frac{2d}{d+2\alpha}} \rceil \leq a^{\frac{2d}{d+2\alpha}} \epsilon^{-\frac{2d}{d+2\alpha}} + 1.$$

Clearly, $x_{1:L} \in \mathcal{Z}_\alpha$ and we obtain

$$\|h - x_{1:L}\|_{\ell_2}^2 \leq \gamma_L^2 \sum_{\ell > L} \gamma_\ell^{-2} h_\ell^2 \leq \gamma_L^2 \|h\|_{\mathcal{Q}_\alpha}^2 \leq \epsilon^2.$$

Consequently, by applying the Hölder inequality for $(\frac{2}{p}, \frac{2}{2-p})$ we obtain

$$\begin{aligned}
\inf_{x \in \mathcal{Z}_\alpha: \|h-x\|_{\ell^2} \leq \epsilon} \|x\|_{\mathcal{Z}_\alpha}^p &\leq \sum_{\ell \leq L} \gamma_\ell^{-p} |h_\ell|^p \\
&\leq L^{1-\frac{p}{2}} \left(\sum_{\ell \leq L} \gamma_\ell^{-2} h_\ell^2 \right)^{\frac{p}{2}} \\
&\leq (a^{\frac{2d}{d+2\alpha}} \epsilon^{-\frac{2d}{d+2\alpha}} + 1)^{1-\frac{p}{2}} a^p \\
&\leq 2^{1-\frac{p}{2}} (a^{\frac{2d}{d+2\alpha}} \epsilon^{-\frac{2d}{d+2\alpha}} \vee 1)^{1-\frac{p}{2}} a^p \\
&\leq 2^{1-\frac{p}{2}} (a^{\frac{2d-pd}{d+2\alpha}} \vee 1) (\epsilon^{-\frac{2d-pd}{d+2\alpha}} \vee 1) a^p \\
&= f(a)g(\epsilon)^{1-\frac{p}{2}}.
\end{aligned}$$

For the second part of the claim, let $\epsilon_n = r_n^{\alpha, \beta, p, q}$ as in Proposition 5.4 and observe that since $n\epsilon_n^2 > 1$ and $p \leq 2$ we have

$$f(n^{\frac{1}{2}}\epsilon_n) = n^{\frac{d+\alpha p}{d+2\alpha}} \epsilon_n^{\frac{2d+2\alpha p}{d+2\alpha}}$$

and since $\epsilon_n < 1$ we have $g(\epsilon_n) = 2\epsilon_n^{-\frac{2d}{d+2\alpha}}$. For $p = 2$ the claim holds trivially since $f(n^{\frac{1}{2}}\epsilon_n) = n\epsilon_n^2$ and $g(\epsilon_n)^{1-\frac{p}{2}} = 1$. For $p \in [1, 2)$, first notice that

$$f(n^{\frac{1}{2}}\epsilon_n)g(\epsilon_n)^{1-\frac{p}{2}} = 2^{1-\frac{p}{2}} \epsilon_n^p n^{\frac{d+p\alpha}{d+2\alpha}},$$

so that $f(n^{\frac{1}{2}}\epsilon_n)g(\epsilon_n)^{1-\frac{p}{2}} \lesssim n\epsilon_n^2$ is equivalent to

$$\epsilon_n \gtrsim n^{-\frac{\alpha}{d+2\alpha}}. \quad (\text{C.25})$$

This is always true, since in the proof of Proposition 5.4 we computed ϵ_n by finding an upper bound, say $B(\epsilon_n)$, on the concentration function (see for example (C.11)), and then solving $B(\epsilon_n) = n\epsilon_n^2$. Since $B(\epsilon_n) \geq \epsilon_n^{-\frac{d}{\alpha}}$, we have that (C.25) holds. \square

Proof of Lemma 5.15. Fix $a, \epsilon > 0$ and let $h \in aB_{\bar{\mathcal{Q}}_\alpha}$. We need to show that \bar{f}, \bar{g} satisfy the bound

$$\inf_{x \in \bar{\mathcal{Z}}_\alpha: \|h-x\|_{\ell^2} \leq \epsilon} \|x\|_{\bar{\mathcal{Z}}_\alpha}^p \leq \bar{f}(a)\bar{g}(\epsilon)^{1-\frac{p}{2}}. \quad (\text{C.26})$$

Noticing that $aB_{\bar{\mathcal{Q}}_\alpha} = \lambda aB_{\mathcal{Q}_\alpha}$ and using Lemma 5.14, we get that

$$\inf_{x \in \bar{\mathcal{Z}}_\alpha: \|h-x\|_{\ell^2} \leq \epsilon} \|x\|_{\bar{\mathcal{Z}}_\alpha}^p = \inf_{x \in \mathcal{Z}: \|h-x\|_{\ell^2} \leq \epsilon} \lambda^{-p} \|x\|_{\mathcal{Z}_\alpha}^p \leq \lambda^{-p} f(\lambda a)g(\epsilon)^{1-\frac{p}{2}} = \bar{f}(a)\bar{g}(\epsilon)^{1-\frac{p}{2}},$$

hence (C.26) is indeed satisfied.

For the second part of the claim, let $\epsilon_n = \bar{r}_n$, for \bar{r}_n as in Proposition 5.8 and observe that since $n\epsilon_n^2 > 1$ and $p \leq 2$ we have $\bar{f}(n^{\frac{1}{2}}\epsilon_n) = \lambda_n^{\frac{(2-p)d}{d+2\alpha}} n^{\frac{d+\alpha p}{d+2\alpha}} \epsilon_n^{\frac{2d+2\alpha p}{d+2\alpha}}$ and since $\epsilon_n < 1$ we have $g(\epsilon_n) = 2\epsilon_n^{-\frac{2d}{d+2\alpha}}$.

For $p = 2$ the claim holds trivially since $\bar{f}(n^{\frac{1}{2}}\epsilon_n) = n\epsilon_n^2$ and $\bar{g}(\epsilon_n)^{1-\frac{p}{2}} = 1$. For $p \in [1, 2)$, first notice that

$$\bar{f}(n^{\frac{1}{2}}\epsilon_n)\bar{g}(\epsilon_n)^{1-\frac{p}{2}} = 2^{1-\frac{p}{2}}\lambda_n^{\frac{(2-p)d}{d+2\alpha}}\epsilon_n^p n^{\frac{d+p\alpha}{d+2\alpha}},$$

so that $\bar{f}(n^{\frac{1}{2}}\epsilon_n)\bar{g}(\epsilon_n)^{1-\frac{p}{2}} \lesssim n\epsilon_n^2$ is equivalent to

$$\epsilon_n \gtrsim n^{-\frac{\alpha}{d+2\alpha}}\lambda_n^{\frac{d}{d+2\alpha}}. \quad (\text{C.27})$$

This is always true, since in the proof of Proposition 5.8 we computed ϵ_n by finding an upper bound, say $B(\epsilon_n, \lambda_n)$, on the concentration function, and then solving $B(\epsilon_n, \lambda_n) = n\epsilon_n^2$. Since $B(\epsilon_n, \lambda_n) \geq \epsilon_n^{-\frac{d}{\alpha}}\lambda_n^{\frac{d}{\alpha}}$, we have that (C.27) holds. \square

D. Proofs of Section 6

Proof of Proposition 6.1. The proof follows the techniques of the proof of [17, Corollary 5] taking into account the form of the Schauder basis functions ψ_{kl} . In particular recall that ψ_{kl} are ϑ -Hölder continuous with $\vartheta \leq S \wedge 1$, $S > \alpha$. Denote by $\kappa_n(W)$ the n th cumulant of a random variable W . Let $u \sim \mu$. Since the odd cumulants of centered random variables are zero and the cumulants are additive for independent random variables, we have for any integer $q \geq 1$ and any $x, y \in [0, 1]$, $x \neq y$

$$\begin{aligned} |\kappa_{2q}(u(x) - u(y))| &= \left| \sum_{k=1}^{\infty} \sum_{l=1}^{2^k} \kappa_{2q}(\xi_{kl}) 2^{-(1+2\alpha)qk} (\psi_{kl}(x) - \psi_{kl}(y))^{2q} \right| \\ &\leq C_q \sum_{k=1}^{\infty} \sum_{l=1}^{2^k} 2^{-(1+2\alpha)qk} |\psi_{kl}(x) - \psi_{kl}(y)|^{2q} \\ &\leq C_q \sum_{k=1}^{\infty} \min \left\{ \sum_{l=1}^{2^k} 2^{2(\vartheta-\alpha)qk} |x - y|^{2q\vartheta}, 2^{-(1+2\alpha)qk} \sum_{l=1}^{2^k} (|\psi_{kl}(x)|^{2q} + |\psi_{kl}(y)|^{2q}) \right\} \\ &\leq C_q \sum_{k=1}^{\infty} \min \left\{ 2^k 2^{2(\vartheta-\alpha)qk} |x - y|^{2q\vartheta}, 2^{-2\alpha qk} \right\} \\ &\leq C_q \sum_{\ell=1}^{\infty} \min \left\{ \ell^{2(\vartheta-\alpha)q} |x - y|^{2q\vartheta}, \ell^{-2\alpha q-1} \right\} \\ &\leq C_q \left(|x - y|^{2q\vartheta} \int_1^{|x-y|^{-\frac{2q\vartheta}{1+2q\vartheta}}} \ell^{2(\vartheta-\alpha)q} d\ell + \int_{|x-y|^{-\frac{2q\vartheta}{1+2q\vartheta}}}^{\infty} \ell^{-2\alpha q-1} d\ell \right) \\ &\leq C_q |x - y|^{\frac{4\alpha\vartheta q^2}{1+2q\vartheta}}, \end{aligned}$$

where C_q is a constant depending only on (q, p) , changing from line to line. For the second inequality we used the ϑ -Hölder continuity of ψ_{kl} as described in

(21) and the bound $(a+b)^{2q} \leq 2^{2q-1}(a^{2q} + b^{2q})$, for any $a, b \in \mathbb{R}$. For the third inequality we used (22), which in turn implies that

$$\left\| \sum_{l=1}^{2^k} |\psi_{kl}| \right\|_{L_\infty} \leq C_2 2^{\frac{k}{2}},$$

by letting $x^* \in \arg \max_x \sum_{l=1}^{2^k} |\psi_{kl}(x)|$ and taking $u_{kl} = \text{sign}\{\psi_{kl}(x^*)\}$.

Since the random variables $u(x)$ are centered, all moments of even order $2q, q \geq 1$, can be written as homogeneous polynomials of the even cumulants of order up to $2q$, hence

$$\mathbb{E}|u(x) - u(y)|^{2q} \leq C_q |x - y|^{\frac{4\alpha\vartheta q^2}{1+2q\vartheta}},$$

uniformly for $x, y \in [0, 1]$. The result follows from Kolmogorov's continuity theorem, since we can choose q arbitrarily large, see [17, Corollary 4]. \square

Proof of Proposition 6.3. As a first step we generalize [51, Lemma 2.1] which holds for standard jointly normal variables, to the case of independent p -exponential variables with $p \in [1, 2]$. Due to independence, we can use the product rule instead of Sidak's inequality. The lemma immediately generalizes due to the estimates in Lemma F.1 below.

To get the result we then follow the proof of [51, Theorem 1.3]. For u drawn from an α -regular p -exponential measure, it holds by (22)

$$\|u\|_{L_\infty} \leq c \sum_{k=0}^{\infty} 2^{\frac{k}{2}} \sup_{1 \leq l \leq 2^k} |u_{kl}| \leq c \sum_{k=0}^{\infty} 2^{-\alpha k} \sup_{1 \leq l \leq 2^k} |\xi_{kl}|.$$

For $\epsilon > 0$, let n be an integer such that

$$\frac{2}{1 - 2^{-\alpha/2}} 2^{-\alpha n} \leq \epsilon < \frac{2}{1 - 2^{-\alpha/2}} 2^{-\alpha(n-1)}.$$

Define

$$b_k := \begin{cases} 2^{\frac{3}{2}(k-n)\alpha}, & \text{if } k < n \\ 2^{\frac{1}{2}(k-n)\alpha}, & \text{if } k \geq n \end{cases}$$

and notice that

$$\sum_{k=0}^{\infty} 2^{-\alpha k} b_k \leq \frac{2}{1 - 2^{-\alpha/2}} 2^{-\alpha n}.$$

Then, if $|\xi_{kl}| \leq b_k$ for all $k \geq 0, 1 \leq l \leq 2^k$, we have that $\|u\|_{L_\infty} \leq \epsilon$. Therefore, by [51, Lemma 2.1] we have

$$\mathbb{P}(\|u\|_{L_\infty} \leq \epsilon) \geq \exp(-C2^n)$$

and the proof is complete since 2^n is of order $\epsilon^{-\frac{1}{\alpha}}$. \square

Proof of Lemma 6.4. Let w_{kl} be the coefficients of w_0 in the wavelet basis ψ_{kl} . Consider $h_{1:K} \in C[0, 1]$ with coefficients $h_{kl} = w_{kl}$ for $k \leq K$ and $1 \leq l \leq 2^k$ and $h_{kl} = 0$ for $k > K$. Then $h_{1:K} \in \mathcal{Z}_\alpha$ for any $K \in \mathbb{N}$ and

$$\|h_{1:K} - w_0\|_{L_\infty} \leq c \sum_{k>K} 2^{\frac{k}{2}} \sup_{1 \leq l \leq 2^k} |w_{kl}| \leq c \|w_0\|_{B_{\infty\infty}^\beta} \sum_{k>K} 2^{-\beta k} \leq c \|w_0\|_{B_{\infty\infty}^\beta} 2^{-\beta K},$$

for $c > 0$ a changing constant independent of w_0 and where for the first inequality we used (22). Choosing $K \in \mathbb{N}$ minimal so that $\epsilon \geq c \|w_0\|_{B_{\infty\infty}^\beta} 2^{-\beta K}$, we get $\|h_{1:K} - w_0\|_{L_\infty} < \epsilon$. Then

$$\begin{aligned} \inf_{h \in \mathcal{Z}_\alpha: \|h - w_0\|_{L_\infty} \leq \epsilon} \|h\|_{\mathcal{Z}_\alpha}^p &\leq \|h_{1:K}\|_{\mathcal{Z}_\alpha}^p = \sum_{k \leq K} \sum_{1 \leq l \leq 2^k} |w_{kl}|^p 2^{p(\frac{1}{2} + \alpha)k} 2^{-\beta k} 2^{\beta k} \\ &\leq \|w_0\|_{B_{\infty\infty}^\beta}^p \sum_{k \leq K} \sum_{1 \leq l \leq 2^k} 2^{p(\alpha - \beta)k} = \|w_0\|_{B_{\infty\infty}^\beta}^p \sum_{k \leq K} 2^{(p(\alpha - \beta) + 1)k}. \end{aligned}$$

The sum on the right hand side converges as $K \rightarrow \infty$ if $\beta > \alpha + \frac{1}{p}$, for $\beta = \alpha + \frac{1}{p}$ blows-up as $K \asymp \log(1/\epsilon)$ and for $\beta < \alpha + \frac{1}{p}$ blows-up as $2^{K((\alpha - \beta)p + 1)} \asymp \epsilon^{\frac{\beta p - \alpha p - 1}{\beta}}$. \square

Proof of Lemma 6.6. Let $h = (h_{kl}) \in aB_{\mathcal{Q}_\alpha}$ and define the approximation $x_{1:K} = \sum_{k \leq K} \sum_{l=1}^{2^k} h_{kl} \psi_{kl}$ for K to be determined below. We have that $x_{1:K} \in \mathcal{Z}_\alpha, \forall K \in \mathbb{N}$ and by (22) and Cauchy-Schwarz inequality, we get the bound

$$\|h - x_{1:K}\|_{L_\infty} \leq c \left(\sum_{k=K+1}^{\infty} 2^{-2\alpha k} \right)^{\frac{1}{2}} \left(\sum_{k=1}^{\infty} 2^{(1+2\alpha)k} \sup_{1 \leq l \leq 2^k} |h_{kl}|^2 \right)^{\frac{1}{2}} \leq c 2^{-\alpha K} \|h\|_{\mathcal{Q}_\alpha},$$

where $c > 0$ is a constant with a value that changes below, dependent only on the Schauder basis, α and later on p . For $K = K(\epsilon; a) \in \mathbb{N}$ minimal such that $\epsilon \geq ca 2^{-\alpha K}$, we have that $\|h - x_{1:K}\|_{L_\infty} \leq \epsilon$. This K satisfies $K = \lceil \log_2(ca^{\frac{1}{\alpha}} \epsilon^{-\frac{1}{\alpha}}) \rceil \leq \log_2(ca^{\frac{1}{\alpha}} \epsilon^{-\frac{1}{\alpha}}) + 1$, hence $2^K \leq 2ca^{\frac{1}{\alpha}} \epsilon^{-\frac{1}{\alpha}}$ and by Hölder inequality we have the bound

$$\inf_{x \in \mathcal{Z}_\alpha: \|h - x\|_{L_\infty} \leq \epsilon} \|x\|_{\mathcal{Z}_\alpha}^p \leq \|x_{1:K}\|_{\mathcal{Z}_\alpha}^p = \sum_{k \leq K} \sum_{l=1}^{2^k} \frac{|h_{kl}|^p}{\gamma_{kl}^p} \leq (2^K - 1)^{1 - \frac{p}{2}} \|h\|_{\mathcal{Q}_\alpha}^p \leq c \epsilon^{-\frac{2-p}{2\alpha}} a^{\frac{2-p+2\alpha p}{2\alpha}}.$$

Therefore, $f(a) = ca^{\frac{2-p+2\alpha p}{2\alpha}}$ and $g(\epsilon) = \epsilon^{-\frac{1}{\alpha}}$ satisfy (6) and the proof is complete. \square

E. Shift spaces of scaled independent product measures

Proposition E.1. *Let ν be the law of the scaled sequence $(\gamma_\ell \xi_\ell)_{\ell \in \mathbb{N}}$, where ξ_ℓ independent and identically distributed univariate random variables and $\gamma = (\gamma_\ell)$ deterministic decaying sequence of positive numbers. Assume that the common*

distribution of ξ_ℓ , has finite Fisher information and variance and has a density ρ_ℓ with respect to the Lebesgue measure which is everywhere positive and continuous. Then for any $h \in \mathbb{R}^\infty$ it holds that the translated measure ν_h and ν are either singular or equivalent. The shift space of the measure ν is

$$\mathcal{Q}(\nu) = \{h \in \mathbb{R}^\infty : \sum_{\ell=1}^{\infty} h_\ell^2 \gamma_\ell^{-2} < \infty\}.$$

Furthermore, letting $\rho_{\ell, h_\ell} = \rho_\ell(\cdot - h_\ell)$, we have for $h \in \mathcal{Q}(\nu)$,

$$\frac{d\nu_h}{d\nu}(u) = \lim_{N \rightarrow \infty} \prod_{\ell=1}^N \frac{d\rho_{\ell, h_\ell}}{d\rho_\ell}(u_\ell) \quad \text{in } L^1(\mathbb{R}^\infty, \nu).$$

Proof. The positivity and continuity assumption on the density of ξ_ℓ , secures that for each ℓ we have that ρ_ℓ and the translate $\rho_{\ell, h_\ell} = \rho_\ell(\cdot - h_\ell)$ are equivalent. Hence by the Kakutani Theorem [6, Theorem 2.12.7] ν and ν_h are either singular or equivalent.

The rest of the proof relies on [31, Section 1] which builds on [49]. In these papers it is shown that if $Z = (Z_1, Z_2, \dots)$ is a sequence of independent of random variables with variance $0 < \sigma_j^2 < \infty$, then a sufficient condition for Z and $Z + \alpha$, where $\alpha = (\alpha_1, \alpha_2, \dots)$, to be singular is that $\sum_{j=1}^{\infty} \alpha_j^2 \sigma_j^{-2} = \infty$. If in addition the Fisher information I_j of Z_j is finite for all j , then a necessary condition for Z and $Z + \alpha$, to be mutually singular is that $\sum_{j=1}^{\infty} \alpha_j^2 I_j = \infty$.

In our assumed setting, since the Fisher information of $\gamma_\ell \xi_\ell$ is $I_\ell = \gamma_\ell^{-2} I$ where I is, the assumed to be finite, Fisher information of ξ_ℓ , and since $\text{Var}(\gamma_\ell \xi_\ell) = \gamma_\ell^2 \text{Var}(\xi_\ell)$, we have that the necessary and sufficient condition for the singularity of ν with ν_h is $\sum_{\ell=1}^{\infty} h_\ell^2 \gamma_\ell^{-2} = \infty$. Since ν and ν_h are either singular or equivalent, the shift space is as claimed.

The Radon-Nikodym derivative follows again from Kakutani theorem in the form presented in [16, Theorem 2.7], noting that in [31, Section 1] it is shown that the Hellinger integral $H(\nu, \nu_h)$ is positive when $\sum_{\ell=1}^{\infty} h_\ell^2 I_\ell < \infty$. \square

F. Estimates for the univariate p -exponential distribution

Lemma F.1. *Let $\xi \sim f_p(x)$, where $f_p(x) \propto \exp(-\frac{|x|^p}{p})$, $x \in \mathbb{R}$, $p \in [1, 2]$. Then there exist constants $0 < r_1 < 1$ and $r_2 > 0$ depending only on p , such that*

$$\mathbb{P}(|\xi| \leq x) \geq \begin{cases} r_1 x, & \text{if } 0 \leq x \leq 1 \\ \exp(-r_2 \exp(-\frac{1}{p} x^p)), & \text{if } x > 1. \end{cases}$$

Proof. For $x \leq 1$ we have

$$\mathbb{P}(|\xi| \leq x) = 2 \int_0^x c_p e^{-\frac{t^p}{p}} dt \geq 2 \int_0^x c_p e^{-1/p} dt \geq r_1 x,$$

where $r_1 = 2c_p e^{-1/p} = \frac{e^{-\frac{1}{p}} p^{-\frac{1}{p}}}{\Gamma(1+\frac{1}{p})} < 1$ for $p \in [1, 2]$.

For $x > 1$, let

$$g_p(x) = \mathbb{P}(|\xi| \leq x) - \exp(-r_2 \exp(-\frac{1}{p}x^p)) = \frac{\int_0^x e^{-\frac{t^p}{p}} dt}{\int_0^\infty e^{-\frac{t^p}{p}} dt} - \exp(-r_2 \exp(-\frac{1}{p}x^p)),$$

for some $r_2 > 0$ large enough, so that $g_p(1) > 0$. Such an r_2 exists since the first term in $g_p(1)$ is fixed and positive and the second term is decreasing to zero as r_2 grows. The derivative of $g_p(x)$ is

$$\frac{d}{dx}g_p(x) = e^{-\frac{x^p}{p}} \left(\frac{1}{\int_0^\infty e^{-\frac{t^p}{p}} dt} - r_2 x^{p-1} \exp(-r_2 \exp(-\frac{1}{p}x^p)) \right).$$

The term inside the parenthesis, as $x \geq 1$ grows, starts from a possibly positive value and is monotonically decreasing, eventually becoming negative. This means that the derivative $\frac{d}{dx}g_p(x)$, as $x \geq 1$ grows starts from a possibly positive value and eventually becomes negative too, and thus has at most one root which corresponds to at most a unique critical point of $g_p(x)$, $x \geq 1$, which if exists is a maximum. Noting that $\lim_{x \rightarrow +\infty} g_p(x) = 0$, and since $g_p(1) > 0$, we get that $g_p(x) \geq 0$, $\forall x > 1$ and the proof is complete. \square

G. Other technical results

Lemma G.1. *Let $q \geq 1$, $d \in \mathbb{N}$ and $\beta > \frac{d}{q} - \frac{d}{2}$. Then $B_q^\beta \subset \ell_2$.*

Proof. Let $w := (w_\ell)_{\ell \in \mathbb{N}} \in B_q^\beta$. For $q = 2$ the claim is trivially true. If $q > 2$, then by Hölder inequality for $(\frac{q}{2}, \frac{q}{q-2})$, we have

$$\|w\|_{\ell_2}^2 = \sum_{\ell=1}^{\infty} w_\ell^2 \ell^{2(\frac{\beta}{d} + \frac{1}{2}) - \frac{2}{q}} \ell^{\frac{2}{q} - 2(\frac{\beta}{d} + \frac{1}{2})} \leq \|w\|_{B_q^\beta}^2 \left(\sum_{\ell=1}^{\infty} \ell^{\frac{2q}{q-2}(\frac{1}{q} - \frac{\beta}{d} - \frac{1}{2})} \right)^{\frac{q-2}{q}},$$

where the last sum is finite if and only if $\beta > \frac{d}{q} - \frac{d}{2}$. If $q < 2$,

$$\|w\|_{\ell_2}^2 = \sum_{\ell=1}^{\infty} w_\ell^2 = \sum_{\ell=1}^{\infty} |w_\ell|^q \ell^{\frac{\beta q}{d} + \frac{q}{2} - 1} \ell^{-\frac{\beta q}{d} - \frac{q}{2} + 1} |w_\ell|^{2-q} \leq c \|w\|_{B_q^\beta}^q,$$

provided $\ell^{-\frac{\beta q}{d} - \frac{q}{2} + 1} |w_\ell|^{2-q} \leq c$. Using that $w \in B_q^\beta$, we have

$$|w_\ell| \leq \|w\|_{B_q^\beta} \ell^{-\frac{\beta}{d} - \frac{1}{2} + \frac{1}{q}}.$$

The last estimate gives

$$\ell^{-\frac{\beta q}{d} - \frac{q}{2} + 1} |w_\ell|^{2-q} \leq \|w\|_{B_q^\beta}^{2-q} \ell^{-\frac{2\beta}{d} - 1 + \frac{2}{q}},$$

which is bounded for $\beta \geq \frac{d}{q} - \frac{d}{2}$. \square

Lemma G.2 (Lemma 3 [40]). *Given $a, b > 0$ consider the functions*

$$r_{a,b}(s) := s^a \log^{-b}(1/s), \quad 0 < s < 1,$$

and

$$v_{a,b}(s) := s^{1/a} \log^{b/a}(1/s^{1/a}), \quad 0 < s < 1.$$

Then we have that

$$\lim_{s \rightarrow 0} \frac{r_{a,b}^{-1}(s)}{v_{a,b}(s)} = 1.$$

H. Minimax and linear minimax rates and their relationship to posterior contraction rates, in the white noise model

Consider the estimation of a function $w \in L_2[0, 1]$ observed under scaled Gaussian white noise, as in Section 4.2. Let d be a metric on $L_2[0, 1]$, and consider the minimax risk over a class $\mathcal{F} \subset L_2[0, 1]$

$$\inf_{\tilde{w}_n} \sup_{w \in \mathcal{F}} \mathbb{E}_{P_w^{(n)}}[d(\tilde{w}_n, w)], \quad (\text{H.28})$$

where the infimum is taken over all estimators \tilde{w}_n constructed using the sample path $X^{(n)} \sim P_w^{(n)}$. The *minimax rate of estimation in d -risk over \mathcal{F}* , is the fastest rate of decay r_n of the minimax risk in (H.28), as $n \rightarrow \infty$. See [27, Definition 6.3.1] for details.

We can also consider a more general minimax framework, and in particular can embed convergence in d -loss in probability in the minimax framework; see [57, Remark (2) in Section 2.1]. In this case, we study the minimax risk

$$\inf_{\tilde{w}_n} \sup_{w \in \mathcal{F}} P_w^{(n)}(d(\tilde{w}_n, w) \geq r_n). \quad (\text{H.29})$$

The *minimax rate of estimation in d -loss in probability over \mathcal{F}* , is the fastest rate r_n for which the minimax risk in (H.29) vanishes as $n \rightarrow \infty$.

In both (H.28) and (H.29), we can restrict the infimum to *linear* estimators, in which case we have the corresponding notions of *linear minimax rates*. The minimax and linear minimax rates in L_2 -risk (as in (H.28)) under Besov smoothness for function estimation in the white noise model can be found in [21, Theorem 1]; see also (14) and (15) in Section 5 of the present article.

The minimax rate in L_2 -loss *in probability* (as in (H.29)) is a benchmark for rates of contraction in L_2 -loss, because if the posterior contracts at a rate ϵ_n at a $w_0 \in L_2[0, 1]$, then the center of the smallest ball containing at least half the posterior mass, is an estimator converging at the same rate ϵ_n in L_2 -distance, in $P_{w_0}^{(n)}$ -probability; see for example [25, Theorem 8.7]. Furthermore, the typical approach for establishing minimax rates in L_2 -risk (more generally d -risk, as in (H.28)) and in particular lower bounds, is by establishing a lower bound in probability and using Markov's inequality to obtain a lower bound in expectation, see for example [27, Section 6.3.1]. Hence, it is implicit in the

typical derivation of the minimax rate in L_2 -risk, that the same rate is also the minimax rate in L_2 -loss in probability. This is indeed the approach used when establishing the minimax rates in L_2 -risk for the white noise model under Besov-type smoothness, see for example [27, Section 6.3.3], and hence the minimax rate m_n defined in (14) is a benchmark for rates of contraction in L_2 -loss under Besov-type smoothness.

We next turn to the question of whether Gaussian priors are fundamentally limited by the minimax rate in L_2 -risk over linear estimators. In the white noise model, Gaussian priors are conjugate to the Gaussian likelihood, hence the posterior is also Gaussian. By Anderson's inequality, see for example Proposition 2.4, the posterior mean which is a linear estimator, coincides with the center of the smallest ball containing at least half the posterior mass. Following the train of thought of the previous paragraph, one would thus expect that the contraction rates in L_2 -loss of Gaussian priors under Besov-type smoothness cannot be faster than the linear minimax rate l_n in L_2 -risk defined in (15). However, linear minimax rates in L_2 -risk over Besov-bodies are established by directly working with L_2 -risk and not by establishing lower bounds in probability, see [21, Section 6] and [22]. In other words, the linear minimax rates in L_2 -risk (as in (H.28)) do not necessarily coincide with the linear minimax rates in L_2 -loss in probability (as in (H.29)), over Besov-bodies. In order to establish that Gaussian priors are fundamentally limited by the linear minimax rate l_n in L_2 -risk, one needs either to establish that l_n is also the linear minimax rate in L_2 -loss in probability (that is, to establish the corresponding lower bound in probability), or to show that, in this Gaussian-conjugate setting, posterior contraction in L_2 -loss at a rate ϵ_n implies convergence in expected L_2 -distance of the posterior mean at the same rate. Both of these tasks appear to be non-trivial.

An alternative, more modest, approach for establishing that posterior contraction rates in L_2 -loss for Gaussian priors under Besov-type smoothness cannot be faster than the linear minimax rate in L_2 -risk, is to study lower bounds on the contraction rate and establish that the upper bounds obtained in this article are sharp. Lower bounds on contraction rates in L_2 -loss for Gaussian priors under *Sobolev* smoothness, have been studied in [12, Theorem 2]. Once more, it is not immediately obvious how this result can be generalized to obtain good lower bounds under Besov-type smoothness, since its proof relies on the weighted ℓ_2 -type structure of Sobolev spaces.

Although beyond the scope of the present paper, we believe that the question of whether Gaussian priors are limited by linear minimax rates in L_2 -risk is extremely interesting; an affirmative answer would rigorously show that when interested in reconstructing spatially inhomogeneous unknown functions (that is functions in B_q^β for $q < 2$), from the point of view of contraction rates in L_2 -loss, it is better to use a Laplace prior rather than a Gaussian prior.